

CUL- H03336-1-0335075

# AMERICAN PHILOSOPHICAL QUARTERLY

①

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson  
Kurt Baier  
Richard B. Brandt  
Roderick M. Chisholm  
L. Jonathan Cohen  
James Collins  
Peter Thomas Geach  
Adolf Grünbaum

Norwood Russell Hanson  
Carl G. Hempel  
Jaakko Hintikka  
Jean Hyppolite  
Raymond Klibansky  
Helmut Kuhn  
Benson Mates  
John A. Passmore  
Günther Patzig

Oliver L. Reiser  
George A. Schrader  
Wilfrid Sellars  
J. J. C. Smart  
Wolfgang Stegmüller  
John Wild  
G. H. von Wright  
John W. Yolton

VOLUME 1/NUMBER 1

JANUARY 1964

## CONTENTS

- |   |    |   |    |
|---|----|---|----|
| I. H. H. PRICE: <i>Appearing and Appearances</i> . . . . .    | 3  | V. GEORGE DICKIE: <i>The Myth of the Aesthetic Attitude</i> . . . . . | 56 |
| II. JUDITH JARVIS THOMSON: <i>Private Languages</i> . . . . . | 20 | VI. BRIAN MEDLIN: <i>The Unexpected Examination</i> . . . . .         | 66 |
| III. B. J. DIGGS: <i>Rules and Utilitarianism</i> . . . . .   | 32 | VII. RICHARD TAYLOR: <i>Deliberation and Foreknowledge</i> . . . . .  | 73 |
| IV. I. THALBERG: <i>Emotion and Thought</i> . . . . .         | 45 |   |    |

# AMERICAN PHILOSOPHICAL QUARTERLY

## POLICY

The *American Philosophical Quarterly* welcomes articles from philosophers of any country on any aspect of philosophy, substantive or historical. However, only original articles will be published, and not news items, book reviews, critical notices, or "discussions."

## SUBSCRIPTIONS

The individual subscription price is five dollars *per annum* for individuals (four dollars for students and for subscribers not resident in the U.S.A. or Canada) and ten dollars *per annum* for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. While available, back issues will be sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscriptions and back orders may be addressed directly to the publisher (University of Pittsburgh Press, University of Pittsburgh, Pittsburgh, Pennsylvania 15213).

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with a wide margin, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively—also being typed with wide margins and spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## REPRINTS

Authors who are subscribers will receive 50 reprints gratis. Additional reprints can be purchased through arrangements made when checking proof.

\* \* \*

The *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.

Application to mail at second-class postage rates is pending at Pittsburgh, Pennsylvania.

335015



# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson  
Kurt Baier  
Richard B. Brandt  
Roderick M. Chisholm  
L. Jonathan Cohen  
James Collins  
Peter Thomas Geach  
Adolf Grünbaum

Norwood Russell Hanson  
Carl G. Hempel  
Jaakko Hintikka  
Jean Hyppolite  
Raymond Klibansky  
Helmut Kuhn  
Benson Mates  
John A. Passmore  
Günther Patzig

Oliver L. Reiser  
George A. Schrader  
Wilfrid Sellars  
J. J. C. Smart  
Wolfgang Stegmüller  
John Wild  
G. H. von Wright  
John W. Yolton



VOLUME 1 (1964)

---

UNIVERSITY OF PITTSBURGH PRESS

# AMERICAN PHILOSOPHICAL QUARTERLY

## CONTENTS OF VOLUME 1 (1964)

	<i>Page</i>
AARSLEFF, HANS <i>Leibnitz on Locke on Language</i> . . . . .	165
ALSTON, WILLIAM P. <i>Linguistic Acts</i> . . . . .	138
BRACKEN, HARRY M. <i>Some Problems of Substance Among the Cartesians</i> . . . . .	129
CHISHOLM, RODERICK M. <i>The Ethics of Requirement</i> . . . . .	147
COBURN, ROBERT C. <i>Recent Work in Metaphysics</i> . . . . .	204
<i>Corrigenda</i> . . . . .	333
DICKIE, GEORGE <i>The Myth of the Aesthetic Attitude</i> . . . . .	56
DIGGS, B. J. <i>Rules and Utilitarianism</i> . . . . .	32
EDIE, JAMES M. <i>Recent Work in Phenomenology</i> . . . . .	115
FITCH, FREDERIC B. <i>A Goedelized Formulation of the Prediction Paradox</i> . . . . .	161
GENTZEN, GERHARD <i>Investigations into Logical Deduction</i> . . . . .	288
HINTIKKA, JAAKKO <i>Aristotle and the "Master Argument" of Diodorus</i> . . . . .	101
ISENBERG, A. I., see T. PENELHUM, W. E. KENNICK, A. I. ISENBERG	
KAHN, CHARLES H. <i>A New Look at Heraclitus</i> . . . . .	189
KENNICK, W. E., see T. PENELHUM, W. E. KENNICK, A. I. ISENBERG	
KYBURG, HENRY E., JR. <i>Recent Work in Inductive Logic</i> . . . . .	249
LEVI, ISAAC AND SIDNEY MORGENBESSER <i>Belief and Disposition</i> . . . . .	221
MEDLIN, BRIAN <i>The Unexpected Examination</i> . . . . .	66
MORGENBESSER, SIDNEY see ISAAC LEVI AND SIDNEY MORGENBESSER	
NELSON, JOHN O. <i>An Examination of D. M. Armstrong's Theory of Perception</i> . . . . .	154
PENELHUM, T., W. E. KENNICK, A. I. ISENBERG <i>Symposium: Pleasure and Falsity</i> . . . . .	81
PRICE, H. H. <i>Appearing and Appearances</i> . . . . .	3
RANKIN, V. K. W. <i>Referential Identifiers</i> . . . . .	233
SCHLESINGER, GEORGE <i>The Problem of Evil and the Problem of Suffering</i> . . . . .	244
SPIEGELBERG, HERBERT <i>Toward a Phenomenology of Experience</i> . . . . .	325
TAYLOR, RICHARD <i>Deliberation and Foreknowledge</i> . . . . .	73
THALBERG, I. <i>Emotion and Thought</i> . . . . .	45
THOMSON, JUDITH JARVIS <i>Private Languages</i> . . . . .	20
WINCH, PETER <i>Understanding a Primitive Society</i> . . . . .	307

### EDITOR'S STATEMENT

The scope of the *American Philosophical Quarterly* is to be the entire range of philosophical inquiry: it will be emphatically synoptic—no branch of philosophy and no approach to the subject will be excluded from the intended purview. The editorial policy is to publish work of high quality regardless of the school of thought from which it derives.

Every issue will be given entirely to articles. Only self-sufficient articles, not news items, book reviews, critical notices, or "discussion-notes" will be published. Contributions may be as short as 2,000 words or as long as 25,000. This policy will make possible the publication of longer and more substantial articles than usually find a place in philosophical journals.

A continuing feature of the *American Philosophical Quarterly* will be the publication of analytical and evaluative surveys of "Recent Work in . . ." various areas of philosophy, taking account not only of books, but of significant discussions in the journal literature. It is hoped that this series will help to keep readers abreast of developments in all regions of philosophy in this day of intensive and far-flung cultivation of our subject.



335075

## I. APPEARING AND APPEARANCES\*

H. H. PRICE

### I. SENSE-DATA, SUBSTANCES AND EVENTS

AS you all know, the sense-datum epistemology is now rejected by the majority of philosophers. There are various reasons for this, and they have various degrees of cogency or persuasiveness. I shall begin by mentioning one of them which I suspect has influenced a good many people, though in a subconscious or unconscious manner, because it is seldom explicitly formulated. It is a metaphysical reason rather than an epistemological one.

We may distinguish between two different ways of looking at the universe, two contrasted metaphysical "world views." On the one hand, you may think of the world in the way Aristotle did, as composed of substances or things, *res per se subsistentes*; and then an event is something which happens to or in a substance, and a particular is either itself a substance or else it is a state of some substance. On the other hand, you may think of the universe as Russell, Whitehead, and the Buddhists do. Then you will hold that events or temporally-brief particulars are the ultimate constituents of the world, and that the notion of substance or thinghood is both imprecise and derivative. On this view what is referred to by a thing-word such as "cat" or "tree" is a complex series of events or temporally-brief particulars, related to each other by relations of temporal and/or spatial continuity and by certain sorts of inductively establishable rules of sequence, sometimes called rules of "immanent causality." Hume's celebrated analysis of personal identity, and the very similar analysis offered by the Buddhists, afford an illustration of this way of looking at the world. So does Whitehead's remark that Cleopatra's Needle is a long event.

Now the sense-datum theory does not fit in at all well with the Substance Metaphysics. Suppose we do accept the Substance Metaphysics, what shall we say about sense-data? What "ontological status" shall we assign to them? A sense-datum, such as a bulgy color expanse or a felt pressure can hardly be a substance, a *res per se subsistens*. It must then be a state of some substance, or an event which

happens to or in some substance. What can this substance be? It cannot be the material object, the thing or substance which we ordinarily suppose ourselves to be seeing or touching. This could only be maintained, if at all, when we are perceiving in absolutely optimal conditions. As soon as the characteristics of the sense-datum diverge ever so little from those we attribute to the object, it becomes impossible to hold that the sense-datum is a state of the object or an event happening to or in the object. For instance, the setting sun sometimes looks oval as it approaches the horizon. It also looks flat; that is, according to the sense-datum philosophers, we sense a visual sense-datum which actually is egg-shaped and flat. We cannot attribute those characteristics to the sun. To what substance, then, can they be attributed?

If sense-data are alleged to be events or particulars existing "on their own," events which do not happen to any substance, particulars which are neither themselves substances nor states of any substance, it will be objected that this suggestion is unintelligible. According to the Substance Metaphysics there just could not be such entities; so all this talk about color-expanes, visual fields, etc., would be sheer nonsense. Alternatively, if there is something which the sense-datum terminology really does apply to, we must be able to find a substance of which these color-expanes, etc., are the states, or to which they happen. And this substance could only be the percipient himself, either his mind or his brain. But that suggestion has the disadvantage of being obviously false. It is just not true that when I see or hear something, what I am aware of is related to my mind in the way thoughts and wishes are, or to my brain in the way that physical or chemical events are. Moreover, if it were true, we should be confronted with the familiar "Iron Curtain" objection. What we are aware of in perception would be just "a private peep show"; and we should have no means of knowing, or even ground for suspecting, that there is an external world at all.

On the other hand, if we reject the Substance Metaphysics and hold instead that the ultimate

\* The gist of this paper was presented at a meeting of the American Philosophical Association (Pacific Division) at Berkeley on December 28, 1962. The paper was written before the publication of the late J. L. Austin's *Sense and Sensibilia* (Oxford, 1962).

constituents of the world are events or temporally-brief particulars, we have no difficulty about fitting sense-data into our metaphysical scheme. We are not, perhaps, compelled to accept the sense-datum epistemology, but at least we have no difficulty in understanding what the sense-datum philosophers are talking about. Moreover the Event Metaphysics enables us to take a much more elastic view of the concept of material thinghood. If a material object is in any case just a very complex assemblage of events or temporally-brief particulars related to each other in various complex ways, it is much easier to suppose that sense-data might be included among the constituents of material objects (or rather that all sense-data except hallucinatory ones might be so included). Such an assemblage of particulars might, for instance, contain members with many different shapes and sizes. Some of the constituents of the sun might be flat and egg-shaped, others flat and circular, while the shapes of still others might be parts of a sphere. Considerations of this kind enable Russell to maintain in *Mysticism and Logic* that sense-data are physical and are included among the "ultimate constituents of matter," and to add later, in the *Analysis of Mind*, that they are mental as well (for surely the same particular might quite well be a member of two different systems or assemblages.)

Whatever we may think of this view of Russell's, I would suggest that the sense-datum epistemology does not fit in at all well with the Substance Metaphysics, whereas it does fit in very well with the metaphysics of events or temporally-brief particulars. And I think there must be some connection between the decline and fall of the sense-datum epistemology and the almost simultaneous decline and fall of the Event Metaphysics. The reasons why the Event Metaphysics itself has been abandoned by most present-day philosophers are not relevant now. I shall only say that they do not seem to me to be conclusive ones. I merely want to suggest that anyone who regards the Event Metaphysics as absurd or unintelligible (as most contemporary British philosophers do) has a strong motive for regarding the sense-datum epistemology as absurd or unintelligible as well.

#### *Sense Data and Appearing*

So much by way of preface. I have mentioned this point about the metaphysical affinities of the sense-datum epistemology because I should like to hear whether any of you think there is anything

in it. But my main object in this paper is a different one. I cannot help thinking that there was an important piece of insight contained in the sense-datum epistemology, whatever mistakes there may have been in the way it is formulated; and it seems to me that that insight is in danger of being lost, now that the sense-datum epistemology has fallen into disrepute. The sense-datum philosophers were passionately interested in the way things *look, sound, feel*, as distinct from the way they physically are. This is what led them to devise a terminology which is neutral as between illusion, hallucination, and normal perception and is designed to bring out the features which are common to all three. I think there *are* such common features, and we are in danger of forgetting it. Moreover, the way things look, as distinct from the way they physically are—including the way they look in highly abnormal conditions—is an important characteristic of the universe as we experience it. It is a fact as solid as any that distant waterfalls look stationary, that the moon looks about as large as the sun and very much larger than any star, that two events which appear simultaneous need not physically be so, that the mountain on the horizon ordinarily looks flat and purplish gray, and that its outline and its internal detail look much less complicated than they physically are.

This interest in the way things *look* is one which the sense-datum philosopher shares with the painter. It could be said, perhaps, that the sense-datum analysis of visual perception is the philosophy of "the painter's eye." Please notice, by the way, that the painter's eye does not see things *flat*, though what it sees does have to be depicted on a flat surface by the painter's hand. On the contrary, anyone who has the painter's eye is particularly interested in that feature of the way things look which is called *visual depth* (I think that some critics of the sense-datum epistemology have entirely ignored what the sense-datum philosophers have to say about visual depth.) Now what the painter sees, or his special way of seeing things, must surely be relevant to philosophical problems about perception. Anyone who wants to give a full account of human sense-experience must surely pay some attention to what the painter's eye reveals. It is not, I think, literally true that the painter sees what other people do not see; that is an extravagant, though no doubt an excusable, way of speaking. But he does have to cultivate a habit of attending to features of the visible world which others are aware of only inattentively.

Russell, in his 1914-20 period, was one of the most ingenious and original of all the sense-datum philosophers. It has always seemed to me that the external world as he describes it in his writings of that period is the painter's world, a world of perspectives or prospects. Or rather, Russell's philosophy in that period is something even more ambitious, an attempt to reconcile or to unify the world as the painter *sees* it with the world as the physicist *thinks* of it. This enterprise of his seems to me one of the most laudable and fascinating that any philosopher has ever undertaken. Unfortunately Russell himself has long since lost interest in it, and so has almost everyone else.

What I have been saying so far amounts to this: The distinctive feature of the sense-datum philosophy is its concern with appearances and appearing. You will remember that Russell, though he was certainly a sense-datum philosopher in his 1914-20 period, often talked in terms of "appearances" instead. I need only quote his dictum, "The thing is the class of its appearances."

## II. TWO SENSES OF "APPEAR"

My next task, then, is to consider the notion of *appearing*. Let us begin with the verb "to appear." The first thing we notice is that this verb is used in two different ways. On the one hand there is the *predicative* use, where the verb "to appear" has to take a predicate: X appears  $\phi$ . Here "appearing  $\phi$ " is the generic concept, of which "looking so and so," "sounding thus and thus," "feeling so and so" (for instance, feeling soft or warm) are the species. We must notice that appearing  $\phi$  does not exclude actually being  $\phi$ . What looks round or feels hot may also actually be round or hot; but equally it may not be. Moreover, "appearing" in this use does not have a contrary; or if it has, the contrary of "A appears  $\phi$ " would be "A is not being perceived at all."

There is also a variant or extension of this use. It might be called the *quasi-existential* use, because it is parallel in some ways to the use of the existential phrase "there is." Here we say, "*There appears to be* a so and so." Thus in describing a mirage we might say "There appears to be a pool of water over there." And if someone were reporting a hallucination, he might say that there appeared to be a tabby cat under the bush, or there appeared to be voices in the next room. After-images, too, are most con-

veniently described in this Existential Appearing language. If you stare at a bright light and then turn away, there appears to you to be a bright patch floating about your field of view and changing color as it does so.

The point here is that in the predicative use, "A appears  $\phi$ ," the speaker is asserting that A does actually exist, though he is neither asserting nor denying that it is actually characterized by  $\phi$ . If he wishes to extend this neutrality to A as well, he will have to say "there appears to be an A characterized by  $\phi$ ."

So much for the predicative use and its extension the quasi-existential use. But there is also another use of the verb "to appear"—the *absolute use*, as one might call it—in which the verb does *not* take a predicate. Is the cat in the house? Well, he appeared at breakfast time, but I have not seen him since. Jones did appear at the garden party, though he only stayed for a few minutes. At 6.25 P.M. the moon began to appear above the horizon. This might also be called the "visibly turn up" sense of the verb "to appear." The cat visibly turned up at breakfast time. Jones presented himself visibly at the garden party. The moon, or the top quarter of it, visibly presented itself at 6.25 P.M.

In this use, "appear" does have a contrary, "disappear." Pussy appeared at breakfast time and then disappeared for the next twelve hours. Jones did appear at the garden party, but when I tried to find him five minutes later he had already disappeared. Moreover, in this use of the word "appear" there is no distinction between what appears and what actually is. If the cat appeared at breakfast time, the possibility that he was not physically there at the time is not left open; it is excluded. Similarly, if Jones appeared at the garden party, it follows that he was actually there, though it may be that he was only there for a short time. But perhaps you may have doubts about some other cases. "Talk of the Devil and he appears." If this is meant literally, would it follow that he is physically present shortly after being talked of? Again in describing a telepathic vision we might say, "Her half brother appeared in the room, though physically he was many hundreds of miles away." Again in the Authorised Version of St. Luke's Gospel, ch. 24, we read "The Lord is risen indeed and hath appeared unto Simon."<sup>1</sup> In such cases as these, you may think, there is no implication that the person appearing was physi-

<sup>1</sup> ὡφθη τῷ Σίμωνι literally "was seen by Simon," but it is interesting that the translators regard "hath appeared to" as equivalent to "was seen by."

cally there. Yes, but I think there is an implication that he actually *was* there in some non-physical or supernatural way which the speaker is not prepared to specify further. In other words, the speaker is denying that the experience was just a visual hallucination. He is claiming that in some sense or other the experience was veridical. (He may of course be mistaken, but that is what he is claiming.)<sup>2</sup>

I spoke just now of the "visibly turn up" sense of the word "appear"; and one might be inclined to say that in this use "appear" is a purely *visual* word, whereas in its predicate-taking use (as also in the quasi-existential use) it covers all the different sense-modalities. But I am not quite sure about this. Suppose you could not see the cat because the lights had fused, but heard a mew in the room; or suppose he suddenly jumped on to your knee while you were sitting there in the dark. I think you might say "Well, well, Pussy has appeared at last"—audibly and tangibly, though not visibly. Again, you would *not* say he had disappeared if you could still hear him purring or mewling somewhere in the room. And you might say "Jones has appeared in court today after all" if you heard and recognized his voice from outside the door, without being able to see him. Again, suppose that something has a very strong and distinctive smell, a skunk for example. I think you might say that a skunk appeared in the garden this afternoon, though you only smelled it and did not see it. You might also say that the smell itself had *disappeared* by now. It would not be very good English to say that a strange smell had *appeared* in the room. But if someone did say this, I think he would be understood by anyone who was not a philosopher.

I conclude then that though "appear" in this absolute or non-predicate-taking sense was doubtless originally a visual word, meaning something like "visibly turn up," it *can* be extended to the other sense-modalities; and in this respect, though not in others, it is parallel to the predicate-taking sense "X appears  $\phi$ ."

Is there also in the absolute sense of the word "appear" a suggestion of suddenness or unexpectedness? Not necessarily of unexpectedness. I expected Jones to appear at the garden party, even though I was not sure just when he would come. In the case of the moon, I expected it to appear above the horizon just about 6.25 P.M., because I had looked up the calendar first or had consulted an astronomical colleague.

But does the verb "appear," in this use of it,

have a continuous present? In the predicate-taking use it has. The distant mountain goes on looking blue for hours. The waiting-room at the railway station felt cold when I entered it, and it continued to feel cold all the time I was there. But could we say "Jones is still appearing at the garden party" if he came ten minutes ago and has not left? Or is appearing in this sensibly-turn-up sense a brief event, rather than a continuing process? Must we say that what continues to be true for a period of time is only the negative proposition "he has *not* disappeared yet"? Well, I think we could perfectly well say, even in the most ordinary English, "He appeared at the garden party for ten minutes and then went away again," however queer it would be to say that throughout those ten minutes he went on appearing there. So there is a sense in which appearing (in this non-predicate-taking use of the word) can be said to go on for a period of time. And though it would be very odd to say that he appeared at the garden party for two hours, it would not be at all odd to say "Professor Schmidt appeared in Oxford for two hours." The point is that the period for a continuous appearing has to be rather short in relation to the normally expected period—the period for which the person or thing in question would normally be expected to be visibly present if he or it visibly turned up there. A guest at a garden party normally stays for half an hour or more; a foreign scholar visiting Oxford normally stays for a day or two. So though an appearing (in this non-predicate-taking sense) may continue through a period of time, the period does have to be a relatively short one—short in comparison with the period for which a thing or person of that particular sort would normally be present and perceptible in that particular environment.

I confess I hardly know how important these rather minute points of ordinary English usage are. I think that the distinction I draw at the beginning between the predicate-taking sense of appearing-verbs (A appears  $\phi$ ) and the absolute or non-predicate-taking sense (A appears) is of some importance and has not always been noticed by philosophers; but some of the other points I have mentioned since then may well seem to you merely trivial. Or perhaps some of you may take the opposite view and may think that I have neglected certain distinctions in ordinary usage which ought to have been discussed. For instance, within the absolute or non-predicate-taking sense of the verb "to appear" should we distinguish between the cases

<sup>2</sup> Perhaps we should make a special heading for this use of the word "appear" and call it the *apparitional* use.

where "X appears" means something like "X visibly turns up," and cases where it merely means "X is visibly (or otherwise sensibly) present," with no suggestion of a sudden arrival or intrusion, such as "turning up" suggests?

Exploring the minutiae of ordinary linguistic usage is an endless task. Once we have embarked on it, we hardly know where to stop, though certainly we must stop somewhere. In this difficulty, the only advice I can offer is that we should remind ourselves of the reasons which make it worth while for an epistemologist to discuss ordinary linguistic usage at all. If our plan is to introduce a technical terminology such as the sense-datum terminology or some variant of it, we do need to build a bridge, so to speak, between our technical terminology and ordinary everyday language. We have to use ordinary everyday language in explaining the meanings of our technical terms. And though we may sometimes help ourselves out by offering *ostensive* definitions, we still have to give directions to the reader for getting himself into the relevant ostensive-definition situation, and those directions themselves have to be given in ordinary everyday language. The term "film-color" will do as an example. To get the reader to understand this technical term of visual phenomenology, we instruct him to look at some fairly distant object through a paper tube, or through a small hole in a door (such as a key-hole). Again, if someone does not understand what we mean by "visual depth" and how visual depth differs from physical distance in the line of vision, we take a photograph with strong contrasts of light and shade and invite him to look at it with one eye shut. (I keep on my mantelpiece a photograph of an ancient Greek statue of an owl which I use for this purpose.)

Or perhaps our plan is to use a nontechnical word like "appear" or one of its derivatives in our epistemological enquiries. But if so, we cannot afford to ignore the points at which we are stretching or modifying the ordinary everyday usage of it; and we must remember that if we stretch it too far we shall not be understood and shall very likely fall into muddles and confusions ourselves. For this reason, we need to pay some attention to what the ordinary everyday usage of the word is, before we start stretching it or modifying it.

But no rigid rule can be laid down to prescribe just how far this preliminary exploration of ordinary usage should go. All we can say is that it should go far enough to ensure that a reasonably intelligent hearer or reader will understand what we are talk-

ing about. And here it is well to remember that the plain man is not nearly so stupid as some modern philosophers think. He can stand quite a lot of stretching of the customary linguistic rules and tumble to your meaning in spite of it. Indeed quite a lot of stretching of them goes on in ordinary intelligent after-dinner conversation, and not merely in the writings of pre-1930 philosophers. Even poets contrive to make themselves understood sometimes.

So much for Ordinary Language. One has to offer one's pinch of incense to this potent and formidable deity which now rules the philosophical world.

### III. TWO PRINCIPLES OF COMMON SENSE PHILOSOPHY

I now wish to say something about two principles which seem to be fundamental in contemporary common sense theories of perception. The first I shall call the Principle of the Primacy of the Normal. The second concerns the use of perceptual verbs such as "see," "hear," etc.; it lays down the rule that these verbs are only to be used in their *veridical* sense. Let us call it the Principle of the Veridicality of Perceptual Verbs. According to this principle it would be a contradiction to say that I saw something which was not physically there or was not physically within the range of my eyesight; or again to say that I heard something which was not an actual physical sound. I have to say something about these two principles because the contemporary common sense philosophers who accept them are the most important and influential critics of the sense-datum epistemology, whereas I wish to suggest that the sense-datum epistemology was not so utterly silly after all, whatever mistakes its exponents may have made.

I begin with the Principle of the Primacy of the Normal. One way of putting it would be this: Illusion and hallucination can only be understood in terms of normal perception; conceptually, they have as it were a parasitic character. Suppose I am deceived by a mirror-image. I believe mistakenly that there is a door in front of me on the far side of the room, but in fact what I see is only a reflection of a door behind me. I could not make this mistake unless I had the *concept* of a door. How did I get this concept? Surely I got it by means of ostensive definition, by hearing the word "door" (or some equivalent word) used in paradigmatic situations. And the same applies to the concept of

"being in front of the speaker." If so, it is impossible that *all* the situations in which someone seems to himself to be seeing a door in front of him should be illusory. There must have been some cases in which there really was a door in front of him; and it must be possible that they should occur again, as in fact they frequently do. If I can have illusions about doors, I must also be capable of applying the concept of "door" correctly, at least in "easy cases." Given reasonably favorable conditions (good light, eyes, and brain in working order, a sufficient degree of attention, etc.) I must be capable of reporting *truly* that there is a door in front of me.

The same applies to hallucinations. Let us take a genuine example, instead of the fictitious one about Macbeth or the dubious one about pink rats, and let us try to describe it in language which will not offend common sense philosophers. Not long ago I experienced a mild visual hallucination myself. It happened in twilight. For a second or two it was as if I were seeing a tabby cat underneath a rather straggly bush; and for a second or two I did actually believe that our own tabby cat, which was missing at the time, had come home again. But in fact there was no cat there. Well, how could I have had this hallucination unless I had known what real tabby cats are like and what it is like to see them?

It may be objected that one can have illusions, and even hallucinations, without being in the least deceived by them, that is, without making any *mistake* at all. This is true, and important too. One of the agreeable properties of the drug mescaline is that it causes non-deceptive visual hallucinations (though in my own case they were only mild and partial ones). One can enjoy them, and one can even describe them in a detached and critical manner while they are actually going on, without in the least taking them for realities. All the same, it can be argued that in order to have an illusion or hallucination one must at least *entertain* a false proposition, even though one need not believe it. To put it another way, the experience is *as if* such and such a material-object proposition were true; although in fact the proposition is false, and one may even know at the time that it is false. The important point is that it is a *material-object* proposition, one into which material-object concepts enter. Unless those concepts are innate ideas, surely one must have acquired them by experience of actual instances, by being confronted with entities which did actually satisfy the concepts in question? And surely if one is to report one's illusory or hallucinatory experience to others, one must

use material-object words to describe it? When I was under the influence of mescaline, the carpet and the surfaces of the chairs and cushions appeared to be heaving up and down very gently, like the waves of the sea on a calm day, and this agreeable experience (which did not deceive me at all) continued for some fifteen minutes. How could I give such a description of this hallucinatory movement, or how could you understand the description, unless both you and I have seen how the sea really does behave on a calm summer day? So the Principle of the Primacy of the Normal still retains its force, even though one may have illusory or hallucinatory experiences without being deceived by them. Such experiences do involve the entertaining of false material object propositions, even though the propositions are not believed; and false material object propositions do enter into the description of such experiences, even though these propositions need not be asserted. It can therefore be argued that we cannot entertain a material object proposition which is false unless we know what it would be like for that proposition to be *true*; and this knowledge can only have been got by perception.

Consequently, the phenomena of illusion and hallucination cannot be used to support a sceptical or radically-agnostic theory of perception, in the way some philosophers have tried to use them. "Since some of our perceptual beliefs are admittedly mistaken, it is at least logically possible that all of them may be mistaken; it is at least logically possible that there are no material objects at all, and even if there are, perception gives us no means of *knowing* that there are any." That is the kind of argument some philosophers have used. On the contrary, it would be said, it is logically *impossible* that all our perceptual beliefs should be mistaken. One cannot make a mistake unless one has the concepts with which to make it; and we could not have any material object concepts at all except by familiarizing ourselves (perceptually) with situations in which those concepts do actually have instances. Thus there would be no abnormal perception unless there is some normal perception: Some perceptual experiences must be what we all in practice think they are—means of getting knowledge (not just questionable beliefs) about material objects and their properties. There would be no illusions and *a fortiori* no hallucinations, unless some perceptions were a means of finding out that there actually are material objects and what they are like.

An important point about philosophical method

follows from this. The sense-datum philosophers laid great stress on the phenomena of illusion, not just for a sceptical purpose, but because they thought that one should begin with these "awkward facts," if one wishes to produce an adequate theory of perception. First find a theory or a formulation, which will fit the awkward cases; and then apply it to the easy or normal ones. Bread and butter first, and cake afterwards. This is what Professor Broad has called the "Principle of Exceptional Cases"; and so far as perception is concerned, it is the exact contrary of the Principle of the Primacy of the Normal which I have been discussing.

Well, suppose we do begin with the awkward or exceptional cases, that is, with the phenomena of illusion and hallucination. Obviously it is no good just talking about them in a general way. One must give concrete illustrations and describe them carefully; and the trouble is that we can only describe them intelligibly if we refer at some point or other to *normal* perceptual situations. When you see a straight stick half immersed in water, it looks bent. How will your reader or hearer manage to *identify* the phenomenon you are trying to describe to him? Obviously, he must know what water is like—*real* water—and what sticks are like, and what it is for one thing to be really (not illusorily) immersed in another. How can he know all this unless some of his perceptions are normal or veridical?

Some philosophers, Prichard for instance, have tried to get out of the difficulty by using the phrase "as we say" or "as we ordinarily say": e.g., "When, as we say, we see a stick half immersed in water. . . ." Doubtless we do ordinarily say such things. But the point is that we say them with understanding, and further, that we know when it is correct to say them. And the situation we are concerned with, in discussing the bent stick illusion, is one in which it is correct to say that we are seeing a stick half-immersed in water.

It will be objected, perhaps, that "correct" here only means "conventionally proper" (as it is "correct" to take off your hat to a female acquaintance when you meet her in the street) and that it does not mean "true." But first, it is difficult to see how there could be conventions at all unless there were a publicly observable inter-subjective world. And secondly, if "correct" here does not mean "true," surely it does mean "near enough to bring true for everyday practical purposes." The old Idealistic metaphysicians used to maintain that all our everyday common sense statements are "in

the last resort" false. But some of them were none the less willing to admit that within the class of everyday common sense statements a distinction must be drawn between those which are *phenomenally* true and those which are *phenomenally* false. Thus (they would say) it is ultimately or in the last resort false to say that I am now sitting in a room in Oxford. For space, time, and matter are not ultimately real. But still the statement that I am now sitting in a room in Oxford is in a way preferable to the statement that I am now riding a bicycle in Birmingham. The second statement is both ultimately false and phenomenally false as well. The first is ultimately false, but it is at any rate phenomenally true. And similarly any philosopher who is talking about illusions has to assume that some perceptual statements are at any rate phenomenally true, if he is to give intelligible descriptions of the kind of situations in which illusions (or hallucinations) occur.

### *The Veridicality of Perceptual Verbs*

We may now turn to the second of the two principles maintained by modern common sense theories of perception, the Principle of the Veridicality of Perceptual Verbs. It is held that the ordinary, and therefore the correct, use of perceptual verbs is what may be called their *veridical* use. If someone sees a motor car coming along the road, this entails that there actually is a motor car coming along the road; and if he hears an explosion, this entails that an explosion does actually occur. To say "I saw it, but it was not actually there" would therefore be a contradiction. And when a mentally diseased person says that he hears voices, at a time when no one is actually speaking in his neighborhood, it cannot be true that he hears them, however sure he may be that he does. (It is logically impossible that he should be hearing them however sure he is that he does.)

Professor Ryle's way of putting this is to say that the words "see," "hear," "smell," etc., are success words, like "win" or "solve" or "find out." This looks like saying that perception is infallible, or that the senses cannot err, the precise contrary of what the sceptics I mentioned before are maintaining. But that would be a misunderstanding. Professor Ryle's point is primarily a linguistic one, about perceptual *words* or about the *expression* "the senses." If, or so far as a mistake is made, what happens must not be *called* perceiving. Instead of calling it a case of seeing or hearing, we must call

it a case of *thinking* that you saw or heard, or *claiming* that you saw or heard. Instead of saying that one of our senses was operating, we must say that it seemed to be operating, but was not. Nevertheless, the point is not wholly a linguistic one. For it is also asserted that cases of seeing, hearing, etc., in this, by definition veridical, sense do in fact occur very frequently; and further, that if they did not occur rather frequently, the material object language could neither be learned nor understood.

The insistence on this linguistic rule or *Diktat* that words like "see" or "hear" are only to be used in their veridical sense is very characteristic of modern common sense philosophers. It is what leads some of them to describe themselves as Naive Realists. I think they regard this linguistic rule as connected in some way with what I have called the Principle of the Primacy of the Normal, or even perhaps as another way of formulating this. I hope that they are mistaken here. I find this linguistic rule about the veridical use of perceptual verbs extremely tiresome and restrictive (for reasons I shall mention presently); whereas the Principle of the Primacy of the Normal seems to me important and instructive, though, as we shall see, it requires some modification (qualifications) before we can accept it. And I think it must be wrong to suggest that the two principles are equivalent or are just two different ways of formulating the same one. The Principle of the Primacy of the Normal does not require us to say that the veridical use of perceptual verbs is the *only* correct use of them, but only that it is *one* correct use of them, which it certainly is; and moreover that there are many situations to which this veridical sense of perceptual verbs does apply, which there certainly are.

Nevertheless, the Principle of the Primacy of the Normal needs to be stated in a carefully guarded way. It is true, and important, that we cannot have an illusion or hallucination without possessing some material-object concepts already. But obviously the concepts in question need not always have been acquired by ostensive definition. We must not forget the old Empiricist doctrine of Complex Ideas. For instance, I have never myself witnessed an earthquake. But if I have heard or read descriptions of earthquakes, I could perfectly well have an hallucination of houses tumbling down, of rumbling noises underfoot, of clocks suddenly stopping, etc. Similarly, I could dream of an earthquake (or perhaps dreaming is to be counted as a special sort of hallucination). It is possible, even, that I might

just have "made up" the concept of an earthquake for myself, if I were a very original person; because my own experience has given me the subordinate concepts (e.g., the concept of a house, the concept of falling down) into which this complex one may be analyzed. Again, I may never have seen a bear, but provided I have read stories about bears—and they need not have been illustrated with pictures—a bush seen in twilight might perfectly well *appear* to me to be a bear.

A more interesting example, because it has a more direct bearing on the sense-datum theory, is this. It is conceivable that someone might never have seen an object which was really bent. Even so, a stick partly immersed in water might still *appear* to him to be bent, provided he had heard and understood some verbal definition of the term "bent," for instance "a line is bent if it is divisible into two lines which run in different directions." Again, suppose you had never seen a biplane (an aircraft with two lifting surfaces one above the other) and I dare say some of you never have; perhaps you have never seen a picture of one. Well, if you suffered from a suitable sort of astigmatism—as I do myself—you might find that a monoplane sometimes *looked* to you to be a biplane, although you had never seen a real biplane in your life, nor even a picture of one. You might even be deceived by this illusion and say to yourself "Now at last I see one of these strange old-fashioned aircraft which I have been told about."

But of course the material object concepts mentioned in these examples—"earthquake," "bent," "bear," and "biplane"—though they need not *themselves* have been acquired by ostensive definition, are nevertheless analyzable in terms of more fundamental material-objects which *have* been acquired in an ostensive manner, by actually observing objects which were instances of them. You may have got your concept of "an earthquake" by means of a verbal definition, or more likely from reading narratives, whether true or fictitious. But in order to get it, you must already have at least the concepts of "earth" or "ground" and of "movement," and these were acquired ostensively, by familiarizing yourself with actual instances of them.

Thus what the Principle of the Primacy of the Normal requires is merely that *some* material object concepts (what one might call the basic or fundamental material object concepts) must be possessed already, before one can have an illusion or hallucination; and from this it does follow that all our

perceptual experiences cannot have been illusory or hallucinatory. But the principle does not require that in order to have an illusion or hallucination of A (say, an earthquake) I must previously have perceived actual instances of A. For A might be a complex concept, and then all that is necessary is that I should have perceived actual instances of the basic material-object concepts into which A may be analyzed.

#### IV. THE MODERN COMMON SENSE ANALYSIS OF APPEARING

We can now return to the concept of *appearing*. The Principle of the Primacy of the Normal and the linguistic rule that perceptual verbs are only to be used in the veridical sense had to be discussed first, if we are to understand what modern common sense philosophers have to say about *appearing*. As we have seen already, the verb "to appear" is used in two distinct ways, one in which it takes a predicate and one in which it does not. We shall confine ourselves to the first for the present. Here the standard formula is "A appears  $\phi$ ," and there are different appearing-verbs for the different sense-modalities, "looking so and so," "sounding so and so," etc. "The mountain looks blue" would be an example. There is also a variant or extension of this use in which "appear" is used in a quasi-existential way: "There appears to be an A characterized by  $\phi$ "; for instance, "There appears to be a pool of water over there which glitters."

The language of "appearing" may seem a promising way of reformulating the sense-datum epistemology. Alternatively, if we wish to retain the sense-datum terminology, the language of "appearing" may seem the most promising bridge for passing over from ordinary everyday speech to the technical phraseology of sense-data (and is in fact used by Professor Broad for introducing his technical term "sensum"). The important point here is that the language of "appearing" is neutral as between illusion, hallucination, and normal or veridical perception. "A appears  $\phi$ " neither entails nor excludes "A actually is  $\phi$ ."

But what about the converse relationship? Obviously "A is  $\phi$ " does not entail "A appears  $\phi$ ." A thing might be round without being seen or touched by anyone; and then it would not appear round, nor would it appear to have any other shape either. Or if it be held that appearing is relative not to a percipient but merely to a point of view, we get into difficulties about concealed objects;

for example, a needle in a haystack, or pencil in the middle of a bag of flour, where there do not seem to be any points of view available from which appearances of these objects could be occurring. The pencil, for instance, would still be cylindrical, but in these circumstances it could not very well appear to be cylindrical nor appear to have any other shape.

But if someone *finds out* by actually perceiving it that A is round, perhaps this does entail that it appears round to him? How else can we find out that it is round unless it looks round to him, or feels round to him when he touches it? No, this is not quite right. There might be a large circular pond. (We shall make it large to get rid of the complication about shape-constancy which arises with small circular objects such as pennies.) Certainly one could find out, by seeing it, that it is circular. But it would only *look* circular if one saw it from directly above, from an aeroplane for instance, and one might have no opportunity of doing this. From all the places from which one was able to see it, it might look (approximately) elliptical. Thus if we try to give a general formulation we shall have to say that when  $\phi$  represents a shape word, finding out by means of vision that A is  $\phi$  entails only that A appears to have some shape projectively related to  $\phi$ , or to have a series of shapes all projectively related to  $\phi$  as we move from one point of view to another; and we shall have to say something analogous where  $\phi$  represents a color-word, such as "blue." This of course corresponds roughly with what the sense-datum philosophers said about the relationship between the physical shape of something and a class or family of many different sensible shapes (and similarly for color).

But the project of reconditioning or rehabilitating the sense-datum epistemology by means of the notion of appearing will not work at all if modern common sense philosophers are right in what *they* say about appearing. They claim to have stopped up that particular bolt-hole, and if the fox cannot escape that way, there is very little hope left for him. That is why I mentioned the Principle of the Primacy of the Normal. According to that principle, illusion and hallucination are as it were parasitic upon normal perception, and it is logically impossible that anyone should be having illusions or hallucinations all the time. Similarly it is argued that "A looks  $\phi$ " can only be understood in terms of "A is  $\phi$ " and is seen to be so (where "see" has the veridical sense mentioned above), and the same view would be held about other appearing verbs

such as "sound" and "feel." But according to the sense-datum philosophy, or any reformulated version of it, one would have to hold that appearing is an essential constituent of *all* perceptual situations, illusory, hallucinatory, and normal. The whole point of the sense-datum terminology is that it is supposed to be neutral as between normal perception, illusion, and hallucination; whatever other differences there may be between them, it was claimed that at any rate the sensing of sense-data was present in all three.

### *Three Uses of Appearing Verbs*

If I understand the modern common sense philosophers rightly, they hold that there are three different senses or uses which appearing-verbs may have. I am going to call them (1) the *inclination to believe* sense, (2) the *probability* sense, (3) the *safety-first* sense. They do not exclude each other, and perhaps we should think of three different nuances or factors within the total sense which an appearing-verb has. But it is convenient to consider them separately.

1. *The Inclination to Believe Sense.* It would of course be absurd to equate "A looks  $\phi$  to X" with "X believes A to be  $\phi$ ," even if we added that he need only believe it very mildly, with nothing like full conviction. Such an analysis would only be plausible if our appearing-verb were being used in a metaphorical or *non-perceptual* sense. ("Jones looks the best candidate to me." Compare "my view is that he is the best," where the perceptual word "view" is similarly used in a non-perceptual sense.) When the sun looks oval to me as it approaches the horizon on a winter evening, I do not for a moment believe that it is oval. I do not even accept this proposition with a mild degree of confidence. I do not accept it at all.

But, it will be said, you do have some *inclination* to accept it, though this inclination is easily and successfully resisted; and that is what you mean, or part of what you mean, by saying "It looks oval, but of course it isn't." If you were less well educated, if you were a simple savage or a child, you might actually believe that the sun does sometimes change its shape on winter evenings and later changes back again.

2. *The Probability Sense.* "A appears  $\phi$ " may also mean "there is some probability (though not certainty) that A is  $\phi$ ." Or perhaps we should call this the *verisimilitude* sense of appearing-verbs. In some

respects the situation is *like* what it would be if the sun really were oval and I saw it to be so. Let us consider another example, the distant waterfall which looks stationary (a pleasing phenomenon sometimes met with in mountainous country).<sup>3</sup> What I see is in some respects like what I should see if there were a thick piece of white wool hanging down the mountainside. We may compare with this the statement "Julius Caesar appears to have been a great orator." Ancient writers do speak of him as if he had been, but we cannot be certain that he was, because none of his speeches have been preserved.

3. *The Safety-first Sense.* This is a kind of *performatory* sense. I say "A appears  $\phi$ " to disclaim the responsibility I should incur if I had said "A is  $\phi$ ." If it turns out later that A is not in fact  $\phi$  (or was not at the time when I spoke) you will not be able to say that I was mistaken and you cannot reproach me for giving you false information. I have insured myself against that retort by using the word "appears" (or some other appearing-verb) instead of the word "is." Nor is it merely a social matter. In somewhat the spirit of Descartes' method of doubt, my appearing-verb insures me against the possibility of having to admit *to myself*, later on, that I was mistaken. Compare again "Julius Caesar appears to have been a great orator." We might say this when we did not wish to take the responsibility of saying that he actually *was* one, or to avoid having to admit to ourselves, later, that we had been mistaken.

\* \* \*

Now I think we must agree that if any of these three ways of analyzing appearing verbs is correct, the Principle of the Primacy of the Normal does apply; it would follow that we cannot understand appearing-expressions unless we understand reality-expressions already. To have even an inclination to believe that A is  $\phi$ , one must know what it is for A to be really  $\phi$ , and one must know what it is for A to be perceived to be  $\phi$ , in the veridical sense of "perceive." And similarly one must already have this knowledge, if one is to understand what is meant by saying that the situation is *like what it would be* if A were actually  $\phi$ , or gives some *probability* (small or great) to the proposition that A is actually  $\phi$ . So too, if I am to disclaim the responsibility of asserting that A is actually  $\phi$  and that I perceive it to be so (in the veridical sense of "perceive"). Thus if these interpretations of appearing

<sup>3</sup> Wordsworth noticed it. Cf. "the stationary blast of waterfalls."

are correct, or if any one of them is, it would follow that the notion of "appearing so and so" is parasitic or derivative. Normal or veridical perception could not be analyzed, not even partially analyzed, in terms of appearing, as the sense-datum epistemology or any variant of it wants to say that it can. On the contrary, appearing would have to be analyzed in terms of normal or veridical perception; "looking blue," for instance, would have no meaning unless we already know the meaning of "being blue and being seen to be so" (in the veridical sense of the word "see").

But are these analyses of appearing correct? I do not deny that appearing verbs are quite often used in each of the three ways mentioned, and sometimes in all three of them at once. But I wish to argue that these analyses omit something. I suspect that in all three of them a familiar metaphorical and *non*-perceptual use of appearing-verbs is taken as the standard use; and it is then assumed that an explanation which works well enough for this metaphorical use will still work when applied to their literal or perceptual use. But to the extent that it does still work, it omits just what is peculiar to the literal and perceptual use of appearing-verbs. We shall be still more liable to overlook this if we use the word "seem" instead of "appear," as some of the common sense philosophers do, or if we treat the two as synonyms. Although the word "seem" is etymologically a visual word, derived from "see" (Cf. the Latin *videri*), it has come to be used most commonly in a *non*-perceptual way; and the three analyses I have mentioned, the inclination-to-believe analysis, the verisimilitude analysis, and the safety-first analysis, do apply rather well to "seeming."

### *The Inclination-to-believe Analysis*

Let us consider the inclination-to-believe analysis. When the sun looks oval on a misty winter day, a few minutes before sunset, I am not sure that I have any inclination at all to *believe* that the sun is now oval in shape or has temporarily become oval. But it could be said, perhaps, that the proposition "the sun is now oval," or the thought of its being temporarily oval, does present itself to my mind and so to speak solicits my belief, even though I have no inclination at all to yield to this solicitation. Very well. But why does *that* particular proposition present itself to my mind at all? The idea of an oval sun, even a temporarily oval sun, is rather odd. Why on earth should it occur to me? And granting

that it does, why should it occur to me at that particular time? The answer is obvious. This very peculiar thought occurs to me because the sun does actually *look* oval in those particular conditions (a rather foggy winter afternoon, shortly before sunset). The sun's looking oval does not *consist* in my entertaining this peculiar proposition about its shape. It is the *cause* of my entertaining that proposition. And if someone in such circumstances really does have an inclination to believe that the sun is now oval, its looking oval does not consist in such an inclination to believe, it is the cause of that inclination. And I suggest that this is always the case when the appearing-verb "look" is used in its primary visual sense. It is not a matter of what you *think*, but of what you *see* or the way you see, in a sense of the word "see" for which the common sense philosophers have not provided. And if the thought (in this case an erroneous one) that A actually *is*  $\phi$  does occur to you, it occurs to you because A does happen to *look*  $\phi$  now, in this primary and purely visual sense of the word "look." If anyone says he does not understand this primary and unmetaphorical sense of the verb "to look," I hardly know how to enlighten him. I can only invite him to put himself into the situation I have described, to turn his gaze toward the sun on a foggy winter evening a few minutes before sunset and attend as carefully as he can to the view which he will then have. I am afraid that this primary and literal sense of the expression "looking so and so" can only be defined ostensively, and the same is true of the primary and literal sense of other appearing-expressions such as "sounding loud" or "feeling hard."

### *The Verisimilitude Analysis*

I now turn to the second analysis I mentioned, the verisimilitude analysis. "The sun looks oval," it is now suggested, means "it is as if the sun were actually oval and I were seeing it to be so" (in the veridical sense of "see"). This analysis is quite correct as far as it goes, but it stops short at the crucial point. In *what respect* is it as if the sun were oval? What feature is there in the situation which makes it like the situation there would be if the sun actually were oval and if I were seeing it to be so? Again, the answer is that here and now the sun does actually *look* oval.

One might put the point rather differently by talking about the experience which the percipient has. His experience is as it would be if he were

seeing an actually oval object in the veridical sense of "see." But again we have to ask *how* or in what respect his experience is like that. And we shall have to answer that the quality "oval" is in some sense presented to him. Or we could say that this quality enters into his experience (not just into his thoughts or the propositions he entertains). It enters into his experience, or has a place in his visual field, in just the same way as the quality "circular" does when he sees the sun in normal conditions, and in just the same way as the quality "oval" itself does when he sees a physically oval-shaped object, an oval-shaped cloud for instance, in normal conditions. This, of course, is very like the sense-datum analysis of "x looks oval"; it is indeed one way in which the sense-datum analysis might be formulated.

In this connection, an interesting point arises about ostensive definition. The situation we are discussing is one which could perfectly well be used for giving an ostensive definition of the word "oval." A man ignorant of Latin might not know that "oval" means, etymologically, "shaped like an egg." Even if he did know this, he might never have seen any eggs or any other things with an egg-like outline. In that case, you might perfectly well teach him the meaning of the word "oval" by drawing his attention to the way the sun looks in the situation I have described. For this phenomenon of the sun's looking oval is of course a *public* one. All the inhabitants of Oxford can see the sun looking oval shortly before sunset on a misty winter afternoon. The phenomenon can also be photographed.

I am tempted to mention another example. There is a familiar color which is actually called "sky-blue"; and though it could, of course, be ostensively defined by means of many other examples, the blue sky is the one we do often actually use for the purpose.<sup>4</sup> Now what sort of an entity is the blue sky? Is it a blue material object, a concave bowl-shaped thing, with a rather "spongy" visible texture and a blue color? Surely not, though it does look like that. Again, when you gaze at the blue sky on a cloudless afternoon, what is it that you are seeing, in the veridical sense of the word "see," the only one the modern common sense philosophers allow? It might be suggested, perhaps, that you are seeing a mass of air particles with dust particles intermingled with them, since that is what there physic-

ally is in the direction you are gazing in. But the common sense philosophers cannot accept this answer. If they did, they would be open to a retort which they themselves have often made (correctly) in criticizing other philosophers. For it would follow that no one ignorant of physics really knows what the phrase "the blue sky" means; and this is absurd, because it is one of the most familiar expressions in the most everyday language, and everyone knows what it means, however uneducated he is. I think a common sense philosopher would be obliged to say that the blue sky is a blue and more or less bowl-shaped entity some miles above our heads. For this, surely, is what the ordinary-language expression "the blue sky" means. But we have the best of reasons for thinking that as a matter of physical fact there is no such bowl-shaped physical entity up above our heads.

The blue sky is rather like the oval-looking sun; again it is rather like a rainbow or a mirage,<sup>5</sup> and it is a little like a mirror-image. The blue sky is a *public appearance*: something which there (publicly) looks to be, but not something which there physically is, though it does of course have a physical cause. Now since this appearing-situation can be used, and indeed actually is used, for giving an ostensive definition of a color-concept, surely we must admit that the color in question is actually being exemplified and moreover actually presented—in fact actually seen, in a sense of the word "see" for which the common sense philosophy does not provide. Or perhaps it would be fairer to say that this very familiar phenomenon presents them with an awkward dilemma. Either they must hold that the veridical sense of the word "see" is not the only one; or else, if they insist that it is the only one, they must admit that what we see is not always or necessarily a material object—that sometimes we see what is not as a matter of physical fact actually there.

### *The "Safety First" Analysis*

What of the third analysis of appearing-verbs, what I called the "safety first" analysis? I say that the distant hillside looks very steep to insure myself against the possibility that it may not be very steep after all. Provided I say "looks" instead of "is," you will not be entitled to reproach me for misleading

<sup>4</sup> Cf. also "horizon-blue."

<sup>5</sup> I am thinking here of the mirages which can be wholly explained in terms of physical optics and are public to everyone in a particular locality, and capable of being photographed. The word "mirage" is also used sometimes (unfortunately, I think) for phenomena which are partly hallucinatory.

you if it turns out later that the slope is quite a gentle one. Here the same sort of question arises as before. Why do I need to insure myself against that particular possibility of error at that particular time? How is it that the idea of the hill's being very steep enters into my mind at all? Because the hill does *look* very steep. Indeed, from this distance, in this clear atmosphere, it looks almost vertical. Thus if my statement has a responsibility-disclaiming character (as it may have when the degree of steepness is practically important, e.g., because we are in a very old motor car which is rather short of oil) this responsibility-disclaiming character is derivative. The appearing-verb "to look" only has this responsibility-disclaiming use because there is a more fundamental and purely descriptive use of the expression "looking so and so" which this "safety first" analysis ignores. The quality of near-verticality is actually exemplified in my present visual field, whether or not it is also exemplified in physical fact. That is what I am telling you when I say that the hillside looks very steep and almost vertical.

#### V. THE BASIC SENSE OF APPEARING-VERBS

So much for the three analyses of appearing-verbs which the common sense philosophers offer us. All three stop short at the crucial point. All three are concerned with secondary or derivative uses of appearing-expressions, and fail to deal with the basic or primary or perceptual use of them.

What we must do is to fix our attention on the basic or primary or perceptual use of appearing-verbs; not on their inclination-to-believe use, or their "it is as if it were" use, or their responsibility-disclaiming use, though these three uses do certainly exist. For instance, we must fix our attention on the literal and visual sense of the expression "looks so and so," the literal and actual use of "this feels hard," "that feels smooth."

It must be admitted that even here there are complexities, especially with the appearing-verbs "to look" and "to sound." Thus ice commonly *looks* cold, walls *look* hard, and distant cumulus clouds *look* bulgy (not flat) when there are strong contrasts of light and shade in them. Roughly, what happens here is that an idea derived from *tactual* experience blends with what we see, and presents itself to the mind as a quality of the seen object in much the same way as the color and shape do. These phenomena—sometimes called phenomena of "complica-

tion"—are interesting and important, but I want to exclude them here because they are not purely visual.<sup>6</sup> *A fortiori* I wish to exclude such examples as "he looks frightened" "that voice sounds English." These are interesting too, but they take us still further away from the purely visual sense of the expression "looks so and so" and the purely auditory sense of the expression "sounds so and so." Moreover, we could not *acquire* our concepts of "cold," "hard," "bulgy" from such experiences as these, for the reason that we must possess these concepts already (at least in some degree) before such experiences can occur. Ice cannot look cold to you unless you are *already* familiar with the quality "cold." The same point applies to such examples as "that voice sounds French" or "this apple smells rotten."

The cases which concern me in this paper are those in which an appearing-situation could be used for giving an ostensive definition of some concept: for instance the sun which looks oval, the distant hillside which looks almost vertical, the room which feels cold, the blue sky which there looks to be (visually appears to be) up above us, the whistling railway engine whose note sounds higher and higher pitched and then lower and lower pitched as the train rushes past. If you have an idea of rehabilitating the sense-datum theory, or of restating in some other terminology the special sort of insight which we think the sense-datum philosophers had, examples of this sort are the most relevant.

Let us consider the sun which looks oval. As I have suggested already, it would be very natural to say that here one's visual field does actually contain something which has an oval shape, something which a painter would depict by putting an oval-shaped patch of reddish-yellow pigment on his canvas. Again, if a photograph were taken of the scene (a stereoscopic photograph if you like) something oval would come out in the top half of the photographic print. And the observer, or anyone else who had been standing near him and looking in approximately the same direction, would say that this oval shape in the photograph was a correct reproduction of something which his visual field had contained at that time. In putting the matter this way, in saying that the visual field did actually contain something with an oval shape, we are obviously coming close to the sense-datum analysis of this appearing-situation.

But what about my suggestion that this appearing-situation could be used for giving an ostensive

<sup>6</sup> If I understand him rightly, Wittgenstein would describe them as cases of "seeing as . . ." We see the lump of ice *as* cold, and the wall *as* hard.

definition of the term "oval"? There is a difficulty about it which does not arise (or not so obviously) about using the blue sky for giving an ostensive definition of the term "sky-blue." Someone may say "How can you get an ostensive definition of the term 'oval' from a situation in which the object is *not* oval? By 'oval' one surely means *physically* oval, and everyone admits that the sun is not physically oval." (A similar difficulty would arise if one tried to define the term "stationary" ostensively by pointing to a distant waterfall which *looks* stationary.)

Well, of course, the sun is *not* physically oval; at any rate we have very good reasons for thinking it is not. But "physically oval" is a much more complex concept than the one which is here ostensively defined. The one which we have ostensively defined is the *simpler* concept of being oval in a plane at right angles to the line of vision. And it seems to me a plain matter of fact that this concept is actually exemplified by something in our visual field when the sun looks oval on a misty winter afternoon a few minutes before sunset. (I mention these circumstances about the time of day and the time of year, in order to give you instructions which will enable you to verify for yourselves what I am asserting.) But this concept—which I am claiming is ostensively definable in the situation described—is only one constituent in the more complex concept of "being *physically* oval." On the other hand, I want to insist that it is a constituent of this more complex concept. Consider an object which is physically oval, a large oval piece of cardboard for example. If it is physically oval, there must be directions from which it *looks* oval in just the way the sun does in the situation described: If you see the object from one of them, the concept of being oval in a plane at right angles to the line of vision is actually exemplified in just the same way as it is now in the case of the sun which appears oval, without physically being so.

To express it more colloquially (though perhaps more clearly) the two things I am asking you to compare are these. I ask you to consider *what you see* of a physically oval piece of cardboard when its oval surface is placed at right angles to your line of vision, and then I ask you to consider *what you see* of the sun, which is not physically oval, in the situation I described. And I suggest that in respect of shape *what you see* of the one object is just like what you see of the other. This not very unfamiliar phrase, "what you see of a so-and-so from such-and-such a place" (where "so-and-so" is a material-object expression and the place and direction are described in

material-object terms), is one which could quite well be used for making the sense-datum terminology intelligible to people who say they cannot understand it. The sense-datum terminology could then be introduced without referring to illusions at all. Illusion, where something appears to be what it physically is not, could be brought in at a later stage. In the language of the sense-datum philosophers, one would begin with normal sense-data, and go on to abnormal ones later.

At any rate, in our discussion of these two examples, the oval-looking sun and the physically oval piece of cardboard, we have clearly been coming closer to the sense-datum analysis; and I think we cannot help coming close to it when we fix our attention on the basic *visual* sense—the primary and literal sense—of the appearing-verb "to look" (and similarly for other appearing-verbs, e.g., "this feels soft"). When something looks  $\phi$  in this literal and visual sense of the word, there is no denying that "looks" comes very close to being a kind of "is." This is shown most clearly when we consider the point I have been trying to make about ostensive definition. When something looks  $\phi$ , we are provided with an ostensive definition of the concept " $\phi$ ." Does it not follow from this that we are being aware of a  $\phi$ -ish particular, an entity which does actually exemplify or instantiate the characteristic  $\phi$ ? And that is very like the sense-datum philosopher's way of treating appearing-verbs. When something looks  $\phi$ , or when there looks to be something  $\phi$ -ish (the blue sky for instance), the sense-datum philosopher would say that we are being directly aware of a particular which actually is  $\phi$ , and this is what he calls a visual sense-datum. And if he were asked what he means by "directly aware of"—a question which has caused a good deal of perplexity—I think he might very well answer by referring to the possibility of ostensive definition. To be directly aware of a  $\phi$ -ish particular, he might say, is to have an experience which could be used for defining " $\phi$ " ostensively.

## VI. PRIVACY AND PUBLICITY

But though this line of thought brings us to something like what the sense-datum philosophers maintain, there is still an important difference about privacy and publicity. I spoke earlier about *public* appearances; and throughout I have been emphasizing the publicity of the phenomena which I have been discussing—the oval-looking sun, the stationary-looking waterfall, the mirage seen on a

hot day; or again, the blue sky and the rainbow, both of which may be used (and are in fact used) for the ostensive definition of color-concepts. On the face of it, nothing which is not public can be used for ostensive definitions. But in most versions of the sense-datum theory a sense-datum is private to the person who senses it; and when you and I both see the sun looking oval, most sense-datum philosophers would say that I am sensing one oval particular and you are sensing another.

Nevertheless this antithesis between privacy and publicity is a more slippery one than is sometimes supposed. Publicity is a matter of degree, not of all or none. The blue sky is public to all normal observers who are on or not too far above the surface of the earth. Yet it does not have the unrestricted publicity which is supposed to characterize material objects. From places high enough up in the stratosphere there is no blue sky. There is a black one instead. And this change does not happen abruptly. As you go higher, there is a gradual transition from bright blue, to dark blue, to blue-black, and finally to black. We might express this by saying that the blue sky is private to a restricted set of points of view, though public as between one point of view and another within that set. These points of view are indeed infinitely numerous (assuming that space is continuous) but they are all contained within a finite region, a region within—say—eight miles from the earth's surface. Similarly the oval appearance of the sun is public to a number of different observers, because it is public to a number of different places or points of view. But here again all these points of view have to lie within a restricted area. They have to be such that there is a large tract of damp and misty air to the west of them (or to the east of them if the phenomenon occurs at or soon after sunrise). Similar restrictions obviously apply to the publicity of rainbows and mirages, to the stationary-looking waterfall, and to the bent-looking stick too, which only looks bent from some directions, but not from all directions.

But what shall we say about the ways of appearing which depend on physiological peculiarities in the observer? Let us consider an astigmatic person who is trying to thread a needle without the help of spectacles or other optical aids. Surely the way the needle and the thread look is private to him? I do not think so, if "private" means absolutely and as it were irremediably private. Physiological optics, as well as physical optics, is a publicly verifiable body of knowledge; and this applies not only to

what physiologists have to say about what goes on in our sense-organs, nerves, and brains, but also to what they say about the *visual* or other sensory phenomena which may be expected to occur when our sense organs, nerves, or brains are in specifiable abnormal conditions. Thus anyone who stares at a bright light for a minute may be expected to have an after-image, and this after-image may be expected to go through a regular series of color-changes. The after-image is private to the person who has it; nevertheless there are a number of propositions about after-images which are publicly verifiable, not merely propositions about the events in the retina which cause them, but about after-images themselves, both their colors and their spatial properties. Again, there is a publicly verifiable *phenomenology* of astigmatic vision, short-sighted vision, hemianopsia, etc., a body of public knowledge with which all oculists have to familiarize themselves for purposes of their professional work. This is a body of public knowledge concerning the way things will *look* to anyone who suffers from any of these physiological defects. These phenomenological propositions have a very considerable degree of publicly-established reliability, so much so that the ways things look, sound, and feel to the patient (or what there appears to him to be which does not physically exist) are very important symptoms for diagnosing the disease which he has. Indeed, medical literature is full of phenomenological propositions; and some of them, be it noted, are phenomenological propositions about hallucinations, phenomena which might be supposed to be wholly and irremediably private, if anything is. Nevertheless, there is quite a large mass of well-established information about the phenomenology of hallucinations; and Dr. J. R. Smythies in a recent book has reproached philosophers for their ignorance of it. (He says, justifiably, that in their discussions of hallucination they depend too much on "hearsay and folk-belief"!)

So if we wish to reformulate the sense-datum epistemology in a way which will do justice to the genuine insight the sense-datum philosophers had, their doctrine about the privacy of sense-data must be viewed with some suspicion. I do not claim to know what the correct analysis of "privacy" is; but I hope I have said enough to show that it is a more complicated concept than it seems, and that the whole antithesis of "private" *versus* "public" needs reconsideration.

Nevertheless, when we consider the way things

<sup>1</sup> J. R. Smythies, *Analysis of Perception* (N.Y., 1956), pp. 2-3.

appear (look, sound, feel, smell) as distinct from the way they physically are, fixing our attention on the literal or perceptual use of appearing-verbs, I think we can see that the sense-datum epistemology was not just the perverse muddle it is sometimes thought to be, whatever defects it may have had. Somehow or other we should try to preserve the spirit of it, whatever changes we may have to make in the letter. And this applies even to hallucinations, which we might be inclined to suppose are *wholly* and irremediably private.

I have not said much about hallucinations in this paper. I should like to conclude by pointing out that the account which the modern common sense philosophers give of them is most unsatisfactory. This will involve some repetition of points I have made before. But at the end of a long and tortuous discussion some repetition may be useful; and what I am going to say will show that if we turn our attention to the phenomena of hallucination, we come close to the sense-datum analysis from another direction.

I have already mentioned the rigid rules which the common sense philosophers wish to impose on us for the use of perceiving verbs such as "see," "hear," "feel" (in the tactual sense of "feel"). They will only allow us to use these verbs in what I called a veridical sense. What one sees must be physically there; otherwise one is not seeing. If you hear something, a bang for instance, it must be a physical occurrence. If you smell a paraffiny smell in the bathroom, there must actually be a paraffiny smell there. If you feel something cold in the basket, there must actually be something cold in the basket.

Now I am by no means sure that verbs like "see" and "hear" are *always* used in this way even by the most ordinary speakers of the most ordinary language, though the veridical use of them is no doubt the commonest. It seems to me quite good English to say that someone saw a ghost or an apparition, that one sees strange landscapes in one's dreams, that some mentally-diseased persons hear voices which are not physically there. And I should expect that the equivalents of these statements are good Dutch too. But I do not lay much stress on this, since I would not like to deny that ordinary ways of speaking (as well as philosophical ones) may sometimes be muddled or confused.

The most important point is that if we restrict ourselves to the veridical sense of perceiving-verbs we have no adequate way of describing hallucinations at all. I shall give a genuine example instead of the fictitious one about Macbeth. At one stage of

my mescaline experience I sat looking at a bed-spread with a rather well-marked pattern on it, and suddenly I had a vivid hallucination of a pile of large and very solid-looking dead leaves, something like holly-leaves in shape, but five or six times as big. The hallucination, which did not deceive me at all, lasted for several minutes. How am I to describe this experience? "You *thought* you saw a pile of large dead leaves." To this I am inclined to reply (too brusquely perhaps) that it was not just a case of thinking, but of *seeing*. I saw those large leaf-like shapes as clearly as I see you now. And if "thought you saw" means "believed that you saw in the veridical sense of 'see'," I can only reply that I believed nothing of the kind. I did not for a moment believe that the leaves were physically there; that was what made the experience so exciting and delightful. Whatever hallucination is, it certainly is not just mistaken belief, and people who talk as if it were can never have had a hallucination themselves. It need not even *result* in a mistaken belief, though sometimes it does.

Moreover, why do you regard it as appropriate to say "I thought I *saw* the leaves" rather than "I thought I felt them with my hands" or "I thought I heard them rustling"? Or again, suppose you tell me that I can say I saw them, provided I put the word "saw" in quotation marks. Why is the verb "see" selected for this honor, or dishonor, of having quotation marks put around it rather than some other perceiving-verb such as "hear" or "feel"? The answer is surely obvious. The experience was *very like* a case of seeing in the veridical sense of "see," whereas it was not at all like a case of hearing or feeling in the veridical sense of "hear" or "feel." Accordingly, some common sense philosophers would encourage me to say "it was *as if* I were seeing a pile of dead leaves," and would encourage a lunatic to say "it is as if I were hearing voices." And I think this formula "it is as if A sees (hears, smells, touches) x" is in fact the best one available to common sense philosophers for describing hallucinations. It is the best that their inhibitions about the use of perceptual verbs leave open to them.

All the same, it is only the best of a bad lot. Of course when someone has a hallucination, it is indeed as if he were seeing or hearing something in the veridical sense of "see" or "hear." The most obvious fact about hallucinations is the resemblance they have to normal or veridical perceptions. But how do they resemble normal perceptions? How or in what respect was my situation like what it would

have been if I had been seeing real dead leaves of an unusual size? Because there was something I actually experienced, something actually present in my visual field along with the walls, floor, etc., which were physically before my eyes; and this something was very like what would have been present if I had been seeing a real pile of large dead leaves, in the veridical sense of the word "see." Given the necessary skill, I could draw a sketch of what there was in my visual field on that occasion, or paint a picture of it. (Blake did in fact paint pictures of his visual hallucinations.) One has to use the word "see" in describing this experience because it was in fact a *visual* experience; and though what was experienced was not physically there, it had color, shape, and location, just as the other parts of the visual field had (the walls, the floor, the legs of the bed which I saw at the same time).

It seems to me that something very like the sense-datum terminology is needed for describing hallucinations. For here one certainly does experience particulars which are not material objects nor parts of material objects; and this time they are private particulars too, though that does not prevent reports about them from being publicly intelligible. Moreover, if my earlier arguments were right, something rather like the sense-datum terminology is also needed for describing the way material objects look, sound, and feel when these appearing verbs are used in their basic or literal sense. And finally, something rather like it is needed for describing cases which are intermediate between ordinary appearing situations and hallucinations,

such as the blue sky or a rainbow or a mirage: entities which are not exactly material objects, for they are not part of our physical environment, and yet they can be completely explained in terms of physical optics, by reference to physical events which are part of our physical environment but are not themselves perceived.

Nevertheless, it might be better to use some other terms instead of "sense-datum" because of the misleading suggestions about infallibility which the word "datum" may arouse. Perhaps it would be better to speak of "appearances," as Russell sometimes did, rather than of sense-data.

Russell's use of the term "appearance" is more closely related to the absolute or non-predicate-taking sense of the verb "to appear" (roughly, the "visibly turn up" sense) than to the predicate-taking sense, in which we say "A appears  $\phi$ ." Perhaps, however, my discussion has incidentally shown that the two senses are not so completely different as we might suppose; for when the sun looks oval, for instance, there is ground for saying that an oval-shaped particular *does* "visibly turn up" since it does occupy a place in our visual field. However that may be, the advantage of talking about appearances, rather than sense-data, is this: Everyone must admit that it is often very difficult to describe appearances, and one may easily make mistakes in trying to describe them. A phenomenological language (and this is what we are looking for) will not be very useful if its rules are such that any statement made in it is bound to be true and even indubitable, provided mere verbal slips are avoided.

*University of Oxford*

## II. PRIVATE LANGUAGES

JUDITH JARVIS THOMSON\*

THERE is a thesis which has been argued for by certain Wittgensteinians, namely the thesis that there can be no such thing as a private language. This thesis is often referred to as, and we are given to understand that it is, both important and original. I hope to show that it is, properly understood, something very familiar and rather trite. But the philosophers who argue for it are no more to be confused with Wittgenstein than the Platonists are to be confused with Plato; so to show this is not to cast discredit on Wittgenstein himself. On the contrary, the disservice was done by those who credited the thesis to him. If nothing else, they failed utterly to take seriously his claim that he held no opinions and put forward no theses in philosophy.

### I

Norman Malcolm<sup>1</sup> says "By a 'private' language is meant one that not merely is not but *cannot* be understood by anyone other than the speaker." This characterization is presumably suggested by *Investigations* #243, where Wittgenstein asks if we can imagine a language of which he says, "Another person cannot understand the language." And a little later on, Malcolm tells us that the "cannot" here is to be a logical "cannot": A private language is one of which it is not merely the case that it is not understood by anyone other than the speaker, but more that it is logically impossible that it should be understood by anyone other than the speaker. The thesis then—and I'll call it Malcolm's form of the thesis, or more briefly Malcolm's thesis—is that there can be no such thing as a language which is private in this sense.

Now what is remarkable is that some philo-

sophers<sup>2</sup> who have understood the thesis in this way nevertheless take it to be relevant to discuss the question whether or not a man who grew up alone on a deserted island could have invented a language for his own use. But why should a man's growing up alone make the language he invents (if he can invent a language) of necessity unintelligible to the anthropologists who discover him? And if what is in question is the possibility of a language which is of necessity unintelligible to anyone else but its speaker, then having shown that a man who grew up alone couldn't have such a language you would still have to do the job of showing that a man who grew up in company couldn't have such a language. Were one to look at the nature of the arguments which have actually been produced for Malcolm's thesis, they would do exactly the same thing in both cases, i.e., the supposed proof of Malcolm's thesis calls for no assumptions whatever about how the supposed speaker of the private language grew up.

It may be objected that Malcolm doesn't *own* the private language thesis, and that there is this or that other thesis (turning on this or that other definition of the expression "private language") which is the *real* private language thesis, and that I should have been dealing with them instead. I'm not going to take time to canvas these other possible theses, for (though I shall not defend this) it seems to me so plain that either they would rule out that English is a language or they rest on the (so far as I can see) utterly unjustified assumption that a man who grew up alone couldn't exhibit behavior which would justify us in saying of him: "Now he is following this rule." So I shall be dealing in what follows strictly with Malcolm's thesis—i.e., the thesis that there can be no private language, where a private language would be one which it was

\* My thanks to James F. Thomson for very helpful comments and criticisms.

<sup>1</sup> "Wittgenstein's *Philosophical Investigations*," *Philosophical Review*, vol. 63 (1954). All further references to Malcolm are to this article.

<sup>2</sup> E.g., Ayer in "Can There be a Private Language?" *Aristotelian Society Proceedings Supplementary Volume* 28. A. J. Ayer says the thesis is, there can be no language which is of necessity unintelligible to anyone but its speaker, and then, oddly enough, thinks he has refuted this thesis by proving (if he has proved even these) the quite irrelevant points (1) that there can be a language which is and always has been as a matter of fact unintelligible to anyone but its speaker, and (2) that sensation-reports are not of necessity unintelligible to anyone but their maker.

Cf. also R. Rhees in "Can There be a Private Language?" *op. cit.*

logically impossible for anyone but its speaker to understand.

## II

But in fact not much has yet been made clear about what the denial of the possibility of a private language amounts to. For the question arises: What could there be about or in a language in virtue of which it would be logically impossible for anyone else but its speaker to understand it?

Wittgenstein says (§258): "Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign 'E' and write this sign in a calendar for every day on which I have the sensation." Now what would make "E" such that it was logically impossible for anyone but LW to know what it meant? For suppose that we were to watch LW very carefully every day for several years, and suppose we were to notice the following regularities in his use of the word "E": Generally when he does write "E" in his diary he has held his cheek and furrowed his brow, he has gingerly touched his tongue to a tooth (and then winced), he has looked into his mouth in a mirror, . . . and he has phoned the dentist. Why are we not entitled to say that by "E" LW means "toothache" or "I have a toothache today"?

Malcolm cites a passage from *Investigations* §243: "E" would be a word in (or itself a small self-contained) private language if it referred to "what can only be known to" LW, to an "immediate private sensation" of his. Now as we described LW's use of the sign "E" it certainly looks as if we do know what the sensation is which he uses "E" to report on; so what we must take to be meant here is this: If in using "E" to report on the having of a sensation LW is referring to that sensation, and if it is the case that no matter how much cheek-clutching and dentist-phoning LW does it is logically impossible that we should know what sensation it is that he refers to by "E," then it is logically impossible that we should know what he means by "E." LW's use of the sign "E" would then be a small self-contained private language.

Who says that when we use sensation-terms (such as "pain," "toothache," etc.), we are referring to our sensations, and that it is impossible that you should know what sensation it is that I refer to when I use these terms of myself? The sceptic about other minds does. Smith alone (he says) can know what it is he's feeling when he thrashes about

like that, or even if he's feeling anything at all.

So the point, then, is this: If the sceptic about other minds is right, then no matter what we observe of LW's behavior, it is logically impossible that we should know what sensation it is that he is feeling, and in particular, what sensation it is that he is referring to by the use of "E." So then it is logically impossible that we should know what he means by "E."

It is worth remarking that if the sceptic about other minds were right, then not only would "E" be a word in (or the whole of) a private language, but so also would *all* the sensation terms. We would each of us be speaking private languages in using the words "toothache," "pain," "hunger," and the rest. And this is as it should be. For one thing which is plain is that the point of raising these questions in this part of the *Investigations* is not to discredit the possibility that you or I should invent a special kind of language, but rather to bring out what queer things would have to be said about our ordinary language if the sceptic about other minds were right.

## III

One last point remains to be cleared up. As I said, the sceptic says that it is logically impossible that anyone but Smith should *know* what Smith is feeling, and it may be argued that this still leaves open the possibility that the rest of us should be entitled on occasion to claim that it is *probable* that he is (say) in pain, or hungry. We should distinguish, then, between Weak Scepticism about other minds, which is the view that we can never know what is in the mind of another, but that we may have good reason for thinking that he is feeling this or that; and Strong Scepticism about other minds, which is the view that we can never know what is in the mind of another, and more, that we are never in a position even to have good reason for thinking he is feeling this or that—that any claim we make about what Smith is feeling can at best be a mere guess. Now it is certainly arguable that Weak Scepticism is not a consistent doctrine—there are very familiar arguments which purport to show that where there can be no knowledge there can be no probability either—and thus that a Weak Sceptic must, to avoid self-contradiction, grow into a Strong Sceptic. I am not going to consider these arguments. I want only to point out that with each degree of Scepticism goes a different strength of "private" in the expression "private

335075

language." With Weak Scepticism goes the view that we can never know what LW means by "E" but that it is conceivable we should have good reason for thinking he means (say) "toothache," and with Strong Scepticism goes the view that we can never know what LW means by "E" and that it is not conceivable that we should have any good reason for thinking he means "toothache" or anything else—any view we have as to what "E" means can at best be a mere guess.

Now I think that what Malcolm means to be denying is the possibility of any strongly private language—i.e., any language in which it is logically impossible we should have any ground at all for supposing we understand. (In fact I think that he would not regard Weak Scepticism as even a consistent view to take, and thus as not interestingly to be distinguished from Strong Scepticism.) I shall not cite the passages which make me think this. It is not necessary, for it would plainly be *harder* to show that there can be no weakly private language than it would be to show that there can be no strongly private language. For by my definitions, if you have proved that there can be no strongly private language, then you would have still to do something more to prove that there can be no weakly private language, whereas proving that there can be no weakly private language calls for proving that which would already make a strongly private language impossible. And this can in any case be seen independently. If I can and could have no reason to think I knew what was meant by "E," then this would surely be a better ground for doubting that "E" was a word in a language than if, while I could never know that I was right, I could at all events form a reasonable conjecture to the effect that it means "toothache." So in attributing to Malcolm the thesis that there can be no strongly private language I do him no injustice: I give him the easier rather than the harder task.

#### IV

The thesis, then, is this: There can be no such thing as a language in which it is logically impossible that anyone else but its speaker should have any good reason at all for thinking he understands. And now why should we suppose that this is true?

There seem to me to be three steps in the argument. The first is this: If a sign which a man uses<sup>3</sup>

is to be a word in a language, his use of it must be governed by a rule or set of rules; and this means that it must be possible for him to use the sign correctly or incorrectly—i.e., on the one hand to follow the rule or act in accordance with it, or on the other hand, to violate or disobey or break it. For R is not a rule unless it is possible either to follow or to violate it.

Straightaway, however, the difficulties rush in. What could *the* or *a* rule governing the use of the words "table" or "pain" or "revolution" be? Indeed it is not only the Wittgensteinians who rely heavily on the notion of a linguistic rule which one could follow or violate, and it is worth bringing out that thus far at all events nothing whatever has been done in the way of making this notion clear. For let us notice that something familiar from Carnap's writings, what he called Semantical Rules, will not do what is wanted here. For example, there is the Semantical Rule that says the sentence "The moon is round" means that or is true if and only if the moon is round; but this isn't anything you could follow or violate because it doesn't say anything about your doing anything. The rules of inference in the more familiar logical systems do at all events say "You may do thus and so"—e.g., the *modus ponens* rule. But what would count as a violation of a rule which says only "You may do thus and so?" What would even count as an instance of following such a rule? Not just doing the thing, for the rule didn't enjoin it on me to do it, but said only that I might do it. It would be a joke to say "I'm only following the rules" as I sit smoking in a carriage marked "Smoking Permitted."

I gather that there are some systems of logic which do contain rules which prohibit—rules which say "You may not do thus and so." These rules could of course be both followed and violated, but it can scarcely be supposed that the linguistic rules meant by the philosophers I have in mind could *all* be of this sort. It would follow that a sign could be a word in a language so long as there was a prohibition against its use.

Malcolm gives an example of a purported linguistic rule: "I will call this feeling 'pain' and will thereafter call the *same* thing 'pain' whenever it occurs." As this sounds more like the adoption of a policy than a rule which I hereby subscribe to, let us rewrite it in the form of a rule—and also replace "table" for "pain" so as not to prejudice

<sup>3</sup> Sometimes the private language argument runs: And what could be the use, the role of the words in a private language? Well, why not simply to keep a record for one's own amusement? And then it seems to me that if this is to be ruled out, the argument will have to proceed as I describe it in what follows.

any issues about private languages: Call this object "table" and thereafter call the same sort of thing "table" whenever—but now whenever what? Whenever you see one? It sounds mad; who would subscribe to a rule which enjoined it on him that every time he sees a table he make some remark about it? But then what should be filled in here? Even "whenever you want to refer to one" would be mad, for why shouldn't I allow myself just to use the word "that" and point to the thing when I want to refer to it? And very much worse still, *what* would I be accepting if that were what I should always do? I am to *call* things "tables," and just what is it to do this? I might say to you "Call Henry," and then if you are obliging you'll call out "Henry!"; but suppose I point to myself and say "Call me Ishmael"—is it so clear what it is I am requesting you to do? Must you ever actually *say* the word "Ishmael"? I'm not saying we don't say this sort of thing, but just that the notion of a linguistic rule is at least as much in need of an explanation as any of the things philosophers make use of it to explain.

But these difficulties may be set aside here, for they are not difficulties peculiar to the private language thesis. Let us set out the form of a linguistic rule for a name "K" of a kind of thing—leaving it open that words which play a different role should be governed by linguistic rules of different forms—as follows: You may call anything of kind X "K" and you may not call "K" anything which is not of kind X. And let us, for the sake of bringing out the rest of the argument, just *pretend* that there are no difficulties with it—we shall pretend that "call" is transparent, and that we need not worry about the fact that one couldn't either follow or violate the first conjunct of the rule. At all events, to subscribe to it is not to commit oneself to always doing something, or to always doing something in this or that kind of situation.

It is to be remarked about rules of this form (and this is an important fact about them from the point of view of the private language argument) that they involve mention of a sign, "K," and of a kind of thing, the X's. What the rule says is that X's and only X's are to be called "K's" and "X" of course can be any expression whatever that stands for a kind of thing.

The first step of the argument, then, can be restated in this way: If a sign "K" which a man uses is to be a name of a kind of thing in a language, his use of it must be governed by a rule of the form, X's and only X's are to be called "K's."

## V

The second step of the private language argument has been thought to be self-evident (so it seems to me) by everyone who has written on this subject at all; and the fact that it is not self-evident is worth bringing out because what it says is peculiar to this argument. What it says is this: A man's use of a sign is not governed by a rule unless it is not merely possible that he should violate the rule but more, that he should violate it unwittingly. That is, it must be possible that he should think he is following the rule and not in fact be following it; from the fact that he thinks he is following the rule, it need not follow that he really is following it. And it is said that this is essential to the concept of a rule. (Cf. Malcolm, p. 532.) Wittgenstein's #202 is generally appealed to here: "to *think* one is obeying a rule is not to obey a rule."

At first sight this second step does look to be self-evident. It seems so clear: After all, if from the fact that I think x-ing is following rule R it follows that it is following rule R, I am then free to do anything or nothing. That "rule" doesn't point in any direction. No restriction is imposed on what I do. R is not a rule but a mere "impression of a rule." (Cf. Malcolm, pp. 536, 537.) Now indeed, a "rule" which rules out nothing isn't a rule. But *does* it follow from the fact that if I think that x-ing is following rule R it is following rule R, that rule R rules out nothing? For why should it not be the case that I think of some activities that if I were to perform them, I would be violating rule R? And then it would certainly seem that if, thinking this of an activity, y-ing, I then y, I *am* violating rule R. The adoption of rule R does seem to impose this restriction on me: I must not y, thinking that to y is to violate rule R.

All this is too abstract. So consider the rule "Always decide to do what you think at the time it would be most fun to do." Note, not "Always do what etc., etc.," for I might think I was following this rule, and not really be following it—e.g., all unknown to me I am paralyzed, and so I am not doing what I think I am doing. Rather, "Always *decide* to do etc., etc." Is there anything in the concept "rule" which rules it out that this should be a rule? Indeed, it even appears that this rule satisfies the first condition on rules, namely that a thing isn't a rule unless I can either follow or violate it. For I could surely, out of some moral compunction, say, decide not to do what I think at the time it would be most fun to do.

It might be objected that it is not at all plain how a man could adopt this rule while not adopting the rule "Always *do* what etc., etc." "All I resolve on is to decide in that way. What I actually do is another matter entirely; I adopt no rule to govern that." This is very odd indeed. "I just do the deciding—which has nothing whatever to do with what I actually *do*." But a man might not have anything nearly so strong as this in mind. One can almost hear a philosopher say to himself, "Now it is not possible to know for certain whether or not one really will be able to bring off what one decides to do. At some times it seems very likely indeed, at other times not. But still, I shouldn't commit myself to anything unless I am absolutely certain I can bring it off. Hence I should adopt 'Always decide,' and not 'Always do.'" (It is reported of Prichard that he would never say "I promise," but only "I fully intend.") Now one might argue that if this is what's behind it, then this man's rule "Always decide" is the same as another man's "Always do"; these are not distinct rules. (Prichard's "I fully intend" was as good as a promise, and was taken for that by his friends.) At all events this calls for an argument—that this is so (*if it is*) is not something that follows from the concept "rule."

Or again, Mummy writes "Whenever you feel the least bit gloomy, think of your Mummy. (It will cheer you up.)" Could I think I was following this rule and not be? (Of course it may be said that it is not possible—knowingly at all events—to violate this rule, so that it does not satisfy the first condition on rules which was set out in step one of the argument. But then surely the possibility of something like this should rather show either that step one of the argument is not true for all rules, or that, while one must be able to violate a rule if it is to be a rule, it is not required that one be able to violate it knowingly. And in fact things are really even worse than this. For notice that one can't even violate this rule unwittingly unless one has, in a sense, forgotten the rule. And now what rules—however private—can't be "violated" in that way? How much help is all this abstract talk about rules anyway?)

To say *S* is not a rule unless I can think I am following *S* and not in fact be, is to set out a condition on rules which would rule it out that rules of the sort I mention here—rules which could in a perfectly ordinary, garden-variety sense be called

private rules—are rules. But that they should be ruled out as not really rules is surely by no means self-evident.

But let us turn to linguistic rules in particular. The form of a linguistic rule for a kind-name "*K*" was to be this: You may call anything of kind *X* "*K*" and you may not call "*K*" anything which is not of kind *X*. So the requirement is this: If the sign "*K*" which a man uses is to be a kind-name in a language, it must be possible for him to think he is following a rule of this form and not in fact be. Now I think it would plainly not satisfy that he might make a slip of the tongue: to be confronted with something which is not *X*, and to mean to say "It's *L*" or "It's *M*," and for "It's *K*" to come out instead. For one thing, one might wish to say that if his utterance of "It's *K*" was a slip of the tongue, then he did not *call* the thing of which he said it a *K*, for this is not what he meant to say of it. For another, it is not plain that following a rule of this kind requires that a man should actually say anything. But more important, since it is always possible that a man should make a slip of the tongue or pen, there would be no sign which a man might use which would be such that his use of it would fail to meet this condition. However private you imagine his use of it, he could think he was saying or writing it and not in fact be. (And moreover, looking ahead to the third step of the argument, that he had made a slip of the tongue or pen could be independently established.<sup>4</sup>) But then if the mistake is not to lie in his thinking he is calling a thing by some other name than "*K*" and its being "*K*" that in fact comes out, it must be that the mistake is to lie in his belief that the thing in question is an *X* when in fact it is not (or vice versa): he calls it "*K*" thinking it is an *X* when it is not an *X*, thus unwittingly violating the prohibition against calling anything which is not of kind *X* "*K*." And so the requirement in the case of linguistic rules for kind-names is the following: If a sign "*K*" which a man uses is to be a kind-name in a language, then it must be possible that he should call a thing a "*K*" thinking it is an *X* when it is not an *X*, where it is the *X*'s and only the *X*'s which (in his use) are to be called "*K*'s."

This certainly *sounds* a reasonable thing to say. But in the first place it is not so far as I can see a consequence of the concept "rule" itself, and in the second place I do not see how it should even be

<sup>4</sup> Unless the sign is "private" too, which is a possibility that no one has remarked on—and this because it has not been noticed that the notion "private language" as Malcolm defines it does not require that the terms of the language be names of sensations. Cf. Section VII below.

examined for correctness or incorrectness until the notion "call" is made clearer to us than has anywhere been done.

But before turning to the third step of the argument there are three further points which should be made:

(1) The Wittgensteinians I refer to—and many other philosophers besides—are in agreement on this, that it is *not* possible that a man should think he is in pain and not be in pain. And it might appear that one who says the following three things contradicts himself—namely (a) I can't think I'm in pain and not be in pain, (b) if a sign "K" which a man uses is to be a kind-name in a language, then it must be possible for him to call a thing a "K" thinking it is of the kind to be called "K" and that it nevertheless not be one, and (c) that "pain" is in fact a word in the English language. In defense of the Wittgensteinians, who do say all these three things, we should bring out that one may say them all without being guilty of an inconsistency. For one thing, it could be said that "pain" just isn't a kind-name—and the Wittgensteinians of course do say this. "Pain" is not the name of a kind of thing (a kind of sensation) but rather something that "replaces pain-behavior." But more important, it could be said that it is not required for "pain" to be a kind-name that a man should be able to think he is in pain and not be, as one can think this is alcohol (from its being a colorless liquid) and it not be (cf. Malcolm, pp. 555, 556). It is only required that he should be able to think that the sensation he now has is of the kind of which he, years ago, thought that sensations of that kind and only sensations of that kind are to be called "pains," and yet that it should not be. (Here "X" in the rule is replaced by "sensation of the same kind as that one.") And this, it might be said, is, for all that has so far been argued at any rate, surely possible—for he might have forgotten just which sensation it was of which he had said to himself (or been told): Sensations of the same kind as this one are to be called "pains." And also the Wittgensteinians could (though they don't) insist that though "pain" is not a kind-name it is all the same a word in the language in virtue of (among other things) its meeting conditions analogous to that which we set out for kind-names—perhaps that a man's use of, e.g., the sentence "I'm in pain" is only a use of a *sentence* in a language if it is possible that he should say "I'm in pain" thinking he is in (roughly) a *situation* (which includes his behavior as well as what is happening to him) of the kind in which he may say "I'm in

pain" and only in which he may say "I'm in pain," and he is not. And a man could make a mistake in this way if he had forgotten just which situation it was of which he had said to himself (or been told): In situations of this kind you may say "I'm in pain." (Malcolm does say [p. 541] that the verbal utterance "I'm in pain," like crying, is in some sense "incorrigible," which if I understand it at all would mean that this condition is not met for the statement "I'm in pain." And it is no wonder Malcolm does this, for he is here trying to derive yet another thesis from the *Investigations*—this one from Wittgenstein's suggestion as to *one* possible answer to the question, how do human beings learn the meaning of the names of sensations. Cf. §244. [And it should be noticed that Wittgenstein nowhere else says anything about "pain" or "I'm in pain" replacing pain-behavior—he says only such things as that these are "tied to" the natural expressions of pain.] But then of course it is equally no wonder that Malcolm has trouble making out how the sentence "I'm in pain" can all the same be used to report on a matter of fact, i.e., be true or false.)

(2) The second point which should be made comes directly out of the first. We said that (for all that has so far been said) one can think something is of the kind to be called a "pain" and it not be; and in exactly the same way it is (for all that has so far been said) possible to think a thing to be of the kind to be called a "decision to do what one thinks at the time it would be most fun to do," or a "gloomy feeling" or "thought of Mummy," and that it not actually be one. But this does not show that it is possible to think one is following the two non-linguistic rules I mentioned above and not be; it does not show the two rules do after all satisfy this second condition on rules. For the two rules did not enjoin it on me that I should or should not, may or may not, *call* anything anything, but only that I form a certain kind of decision, think a certain thought. And it is not to be understood that one who grants "It is not possible to think one is in pain and not be" should also insist "But it is possible to think one is making a certain decision, feeling gloomy, thinking of Mummy, and not be."

And lastly, (3), it may be thought that this condition which a man's use of a sign "K" must satisfy if it is to be a kind-name in a language *already* rules it out that a private language should be a language, and we should bring out that it does not. Let us go back to LW who invented the sign "E" to report on the recurrence of a certain kind of

sensation. Wittgenstein said (#258): "Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign 'E' and write this sign in a calendar for every day on which I have the sensation." And it may be asked: How does the sign *get* associated with a kind of sensation? Presumably LW gives himself a "private ostensive definition." He fixes his attention on a sensation, and says to himself: You may call this sensation and any others of the same kind as this one "E." Same kind? *What* kind? Supposedly "sensation of the same kind as *this* one" is to stand for a kind of sensation, but how do I get it to do that? Just by fixing my attention on this sensation? For there to be a kind of sensation (kind of thing in general) there must be limits as to what is or isn't of that kind; and how should I get limits to the kind marked out by simply fixing my attention on one given sensation? But if I don't identify a kind of sensation, then the rule I recite to myself is mere words—my future conduct is not directed in any way.

The question, "How do I identify a kind of sensation?" is a very respectable philosophical question. But of course it is only a special case of the very respectable philosophical question, "How do I identify a kind of *thing*?" And we may ask: Just why is there supposed to be any more difficulty about a private ostensive definition of a newly invented sensation-word than there is about a public ostensive definition of a newly invented thing word? *Why* any more trouble about identifying a kind of sensation than about identifying a kind of *thing*? Suppose I were to set about introducing by an ostensive definition a new name for a shade of red or blue I would like to identify or (for a particularly relevant example) for a kind of squiggle I would like to identify. I write a capital letter E on a bit of paper, and I say, that and any others of the same kind are to be called "E's." Have I identified a kind of squiggle? Well, what other marks would also be of the same kind? Would an italicized capital E be of the same kind? Or one with a long middle bar? Suppose when asked I said, "Yes they are of the same kind," or, "No, they are not of the same kind." Need I have had them specifically in my mind when I introduced the kind-name "E" in order for it to be the case that

I was indeed picking out a kind of squiggle? Surely not.<sup>6</sup> Or better, if anything of this sort *were* necessary if I were to have identified a kind of thing, then it could have been present when I identified a kind of sensation—i.e., if I could have had other squiggles in mind as also of the same kind, then I could have had other sensations in mind as also of the same kind.

Indeed, private ostensive definition of a newly-invented sensation-word is supposed to be in a worse position than public ostensive definition of a newly-invented thing-word. But then something else will be required to show that it is worse than just this; that it is a matter of fixing one's attention on a sensation and saying of it: Any sensation of this kind is to be called "E" or "K" or "pain." It would have to be shown why its having been a *sensation* (an "immediate private sensation") that I fastened my attention on made it impossible that I should in this way identify a kind, whereas its having been a thing that I fastened my attention on did not make it impossible that I should in this way identify a kind. And in fact the third step of the argument can be regarded as an attempt to make just this very point.

## VI

Now this is where we have come so far. The thesis that there can be no private language is the thesis that there can be no language which it is logically impossible that anyone but the speaker should understand. And the argument for this thesis has so far proceeded in the following way: (1) It was said that if a sign which a man uses is to count as a word in a language, his use of it must be governed by a rule—here specifically, if a sign which a man uses is to count as a kind-name in a language, his use of it must be governed by a rule of the following sort: You may call anything of a kind X "K," and you may not call anything "K" which is not of kind X. (2) If a sign which a man uses is to be governed by a rule of this sort it must be possible that he should call a thing a "K" thinking it is of the kind to be called a "K" and it not be.

The third step of the argument is this: There is no such thing as a man's thinking a thing is of the kind to be called "K" and it not being so unless it is logically possible that it be *found out* that it is not

<sup>6</sup> Does the child have to write "dog" just exactly as you do? Anything within a certain range will do. Of course in identifying for him a class of "dog"-inscriptions, you weren't picking out a new class of thing (you were taught what the class is just as he is now). All the same, you pick out a class for him in writing "dog," which he may learn from one sample alone. Why should I not be able to form a new class—of squiggles—and pick it out for you in the same way?

so. And in the positive: There is no such thing as a man's thinking a thing is not of the kind to be called "K", and it being so unless it is logically possible that it be *found out* that it is so.

Since it is so plain when this third step is set out baldly like this what it is all going to amount to, it is worth stressing that this third step of the argument is required if the desired conclusion is to come out of it. That it is involved in the argument as actually set out might be denied, for it is nowhere set out explicitly. (Though I think it would be difficult to deny this. Consider, for example, Malcolm's queries: "Now how is it to be decided whether I have used the word consistently? What will be the difference between my having used it consistently and its *seeming* to me that I have? Or has this distinction vanished?" Of course one doesn't know for certain what a philosopher is appealing to when his argument for a thesis proceeds by the method of rhetorical questions, but it seems to me that unless Malcolm is taking it that, where it is not possible to "decide" whether I have used the word consistently the distinction between my having used it consistently and its seeming to me that I have *has* vanished, then this passage is simply incomprehensible. But we may in fact appeal to a more explicit passage, which follows shortly after: "My impression that I follow a rule does not confirm that I follow the rule, unless there can be something that will prove my impression correct." And then follows Wittgenstein's ironic comparison with the man who assures himself that what was said in his copy of the morning paper was true by buying several more copies of the same paper.<sup>6</sup>) But whether or not it could be denied that this claim was in fact being made by those who have argued against the possibility of a private language, we should stress this: That *without* this claim, the argument simply stops dead in its tracks, since nothing that has so far been said rules out the possibility of a private language. LW might say: "Indeed, there must *be* a difference between my sensation's being of the kind I had decided to call 'E' and its seeming so to me if 'E' is to be a kind-name of a sensation; but why should it follow from the fact that my sensations are strongly private that there is no such difference?" It might be said: "Well, how is it to be established that the sensation is or is not of the

required kind?" And now if the argument is to proceed, there must be something which rules it out that LW should quite acceptably reply: Perhaps it can't be found out that my sensation is or is not of the required kind, but all the same it may be that it is.

And what rules this out is what I have called the third step, which simply amounts to the denial that there is any such thing as a thing's being or not being of the kind to be called "K" or "E" over and above its seeming to a man that it is unless it is logically possible that it be found out that it is or is not of that kind. I say, "Logically possible that it be found out"—for I think it cannot be supposed that what used to be called the "technical" possibility of finding out is required. (Remember how careful the Positivists were in regard to all these distinctions?) For if more than the mere logical possibility is required, then far more is going to be ruled out than just private languages—e.g., it would be ruled out that "ace" was a kind-name in English on the ground that I have just destroyed a playing card without looking at it, and so it is not any longer as a matter of fact possible to find out whether or not it was an ace. But of course you would always need a stronger premise than just that something *is* the case, e.g., "I destroyed a playing card without looking at it," in order to get as a conclusion that it is not logically possible that something should be the case, e.g., "It is not logically possible that we should find out whether or not the card was an Ace." (Or rather, to be more precise, there is no familiar rule of inference which will take you from a contingent premise to the conclusion "It is logically impossible that . . ."—except the rule of inference which allows you to derive a necessary truth from *any* proposition whatever. And then of course you would have already to know independently that the conclusion "It is logically impossible that . . ." is a necessary truth.) As far as the playing card is concerned, it is surely logically possible that, e.g., someone else had looked at it before I destroyed it.

In fact I think (though I shall not try to show this here) that only a private language will fail to meet this condition. Any kind-name "P" such that there is one occasion on which it is not logically possible to find out whether or not "P" applies to a thing must necessarily be a private kind-name—

<sup>6</sup> Cf. Mr. R. Rhees (op. cit.): "And if every use of the mark is also a definition—if there is no way of discovering that I am wrong, in fact no sense in suggesting that I might be wrong—then it does not matter what mark I use or when I use it." Unless I am muddled as to syntax, the "in fact" does a job here which might equally well have been done by "or, which amounts to the same."

i.e., one such that it is not logically possible to find out on any occasion of its use whether or not "P" does apply to a thing.

But what we should show is that this is true of private kind-names; what we should show is that if "E" is a word in or the whole of a private language, then it is not anywhere at any time for any person logically possible to find out that "E" applies or is true. And thus that, whether or not anything else does, at all events every sign in a private language *does* fail to meet the condition for being a kind-name in a language laid down in this third step of the argument.

To bring this out we have to fill in a certain vagueness in the statement of the condition. I put the condition in the words "unless it is logically possible that it be found out that a given thing is or is not of the relevant kind," and the question arises what "finding out" is to mean here. Must it be logically possible to establish conclusively that the thing is or is not of the relevant kind? Or would it be enough that it was logically possible that one should have or obtain good reason for thinking the thing was or was not of the relevant kind? (Malcolm does use the word "prove" here—I quoted the passage above—but in fact the Wittgensteinians use the word "prove" and the related word "criterion" in so relaxed a manner that it is not really fair that we should take the appearance of this word as proof that it is "proof" that is meant.) It seems to me that it is very likely that it is the former that is required—as we might call it, strong finding out. For one thing it appears to me that the two possibilities would not be regarded as distinct—we have already mentioned the familiar arguments purporting to show that there can be no probabilities where there can be no knowledge. (Cf. above on the distinction between Weak and Strong Scepticism.) But more important, it is not plain how a private language would be ruled out if it were not strong finding out what is required. For consider LW again: Could it not happen that he should think his present sensation was of the kind to be called "E"—and then later think "Oh what a fool I am. Now I remember what the sensation was which struck me that afternoon last May. And I was wrong just now in thinking the one I just had was of that kind—it wasn't like it at all." It could of course be said: This appealing to further impressions shows nothing unless it produces an impression "which is actually correct" (cf. Malcolm, pp. 533, 534; on memories). But to take this line suggests that it is conclusive or strong

finding out which is required—either because good reasons are not enough, or because where there can be no strong finding out there can be no good reasons either. A case could *perhaps* be made for saying: This that happens later is not a reason for thinking that his former impression was false, but just itself the impression that his former impression was false. A good reason for thinking the former impression was false would be enough—strong finding out is not necessary—but the impression that you were or were not mistaken cannot amount to a reason for thinking it so. However, this is not at all clear; and it may in fact be absurd.

I am inclined to think that the finding out which is mentioned in the third step must be a conclusive or strong finding out if private languages are to be ruled out. But let us leave all this open. At all events, the possibility LW may later have the impression he was mistaken is not to count as the possibility that he should find out he was mistaken.

People have raised the question in connection with the denial of the possibility of a private language just why "corrigibility by others" should be required if the sign a man uses is to be a word in a language. So far as I can see it is not specially corrigibility by others that is being required, but rather just corrigibility. Corrigibility by others comes in in this way. The Strong Sceptic about other minds says that no amount of observation of LW's behavior could give us any reason whatever for thinking we knew what sensations he was having at any time—thus we are not in a position to find out whether or not he makes a mistake in thinking the sensation he now has is of a kind which he formerly had. But then similarly no amount of observation of LW's behavior by LW himself could give *him* any reason for thinking he knew what sensations he was having at any time—and in particular, no behavioral traces (such as a photograph of himself he had taken the day he was first struck by the sensation he wished to record) could give him any more reason for thinking he knows what he then felt than it could give us reason for thinking we know what he then felt. So what makes it impossible for us to correct his use of the sign "E" goes halfway to making it impossible for him to correct his use of the sign "E"—and if there were that which would make his use corrigible by us, it would also make his use corrigible by him. But it is supposed by hypothesis that there isn't that which would do this. Now add to the strong privacy of the sensations whatever it is that you need to make out

in addition that LW's later impressions that he was or wasn't mistaken can't count as LW's finding out that he was, and you now have it that no one can find out whether or not he is now mistaken.

A private language, then, will fail to meet this condition, so long as we are given either an appropriate account of "finding out" (i.e., strong finding out) or an appropriate account of "later memory impressions" (i.e., that they could not count as finding out).

Well, is this third step of the argument true? I shall not bother to bring out the things that can be said in defense of it, or the things that can be said against it, for it will be plain what sorts of things they are the moment we conjoin the three steps of the argument to form one general principle, which is what I want to say the denial of the possibility of private languages amounts to. The general principle is this: A sign "K" which a man uses is not a kind-name in a language unless (by 1) he has identified a kind of thing to be called "K," which will only be the case (by 2) if it is possible for him to call a thing a "K" thinking it is of that kind when it is not, which will only be the case (by 3) if it is possible to find out whether the thing is or is not of that kind. Or in sum: A sign "K" is not a kind-name in a man's language unless it is possible to find out whether or not a thing is of the kind associated with "K" (over and above its seeming or not seeming to him to be so)—or more simply still, a sign "K" is not a kind-name in a man's language unless it is possible to find out whether or not a thing is a K.

And it is plain that this is nothing more than a revised formulation of something very familiar indeed, namely the Principle of Verification.<sup>7</sup> We are no longer to say that what purports to be a kind-name "K" has meaning if and only if it is possible to find out whether or not a thing is a K. But we are instead to say that what purports to be a kind-name "K" is a kind-name in a man's language only if it is possible to find out whether or not a thing is a K. Strong finding out goes with what used to be called Strong Verificationism; weak finding out goes with what used to be called Weak Verificationism. And the change, then, amounts to this: from "if and only if" to "only if"; and from "is meaningless" to "is not a kind-name in a man's language." Perhaps the reasons why this might

not be thought to be much of an improvement will not need to be set forth.

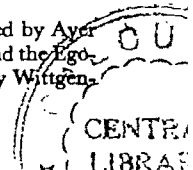
Since the denial of the possibility of a private language is surely a denial of the possibility of a pure sense-datum language—since the thesis we are concerned with is often appealed to in order to discredit phenomenalism—it may seem strange to suggest that the private language thesis is really a mere restatement of verificationism. After all, in recent philosophical history, verificationism has been taken to lead directly to phenomenalism. But we have only to reflect that there were physicalists as well as phenomenologists who rested their case on verificationism. Verificationism does not lead by itself either to one or to the other; your choice between them rested not on the question whether or not you adopted the Principle of Verification, but rather on the question what kinds of ascriptions you took it to be possible to verify. And on this matter, our Wittgensteinians are at one with the physicalists.

## VII

It may be asked, how can it be that the thesis we are considering really amounts to something quite general, like a revised formulation of the Principle of Verification, when all that was at stake was a private language—a language which a man invents to report on his own sensations, or the language which we would be using in reporting on our sensations if the sceptic about other minds were right.

In fact, if we take seriously Malcolm's definition of the expression "private language," then it is not merely our use of sensation-terms which (if the sceptic about other minds were right) would count as private languages, but many other things as well. Malcolm said that a private language would be (if there could be such a thing) "one that not merely is not but [of necessity] *cannot* be understood by anyone other than the speaker." I now give a recipe for constructing hypothetical languages which would be private in Malcolm's sense. Take any classical metaphysical problem on which there is such a thing as a Sceptical View. Now take the range of statements of which it is the relevant Sceptical View that we can never know whether or not any statement in that range is true. And now

<sup>7</sup> That there is a connection between the private language argument and verificationism has already been suggested by Ayer (op. cit.) and Strawson (reference given below). The connection is also noticed by Carl Wellman, in "Wittgenstein and the Ego-centric Predicament," *Mind*, vol. 68 (1959), April number—though this article seems to me to underestimate not only Wittgenstein, but also the force of the private language argument.



if the sceptic's view were correct, then what we might call the relevant term or kind of term or form of term will be such that if a man uses that term or kind of term or form of term, he speaks a (at least partly) private language. Consider the metaphysical problem about the existence of material objects. And now suppose that the sceptic about material objects were right—i.e., that it is not possible that any of us should know of any term purporting to stand for a material object whether or not it really applies to anything. So, in particular, we are supposing that it is not possible that anyone should know of the word "table" whether or not it really applies to anything. We are supposing that although it can *seem* to us as if this really is a table, it is not possible that anyone should *know* that it is one (Weak Scepticism), nor even possible that anyone should have good reason for thinking it is one (Strong Scepticism). Now suppose some man—let us arbitrarily call him Moore—insists that *he does* know of many things that they are (and not merely appear to be) tables. "Here's one, for example," he says, pointing. But now what shall we take Moore to be saying here? That it looks as if he's pointing to a table, that generally when he uses these words there appears to be a table in the vicinity, is no proof (Weak Scepticism), or even good reason (Strong Scepticism), for thinking that he is pointing to a table, that generally when he uses these words there is a table in the vicinity. So how should we know—suppose ourselves to have good reason for thinking we know—what he means to be saying is here when he says "Here's a table"?

All that is needed for turning a class of sentences "S" into sentences of a private language is that whatever it is which is a way of finding out whether or not sentences of kind S are true should be made logically irrelevant to the truth of sentences of kind S. *Instantly* it is not possible that we should know whether or not any sentences of kind S are true; and *Instantly* if any man uses sentences of kind S it is not possible that we should know what he means by them.

Indeed it is surprising that those who have tried to derive the thesis we have been considering from the *Investigations* have failed to notice this possibility of generalizing the notion "private language."<sup>8</sup> Particularly in view of the fact that, as Malcolm

notes, Wittgenstein explicitly suggests (cf. e.g. p. 180) that the relation between reports about the behavior of another and reports about the contents of the mind of another is analogous to the relation between reports about one's sense-impressions and reports about physical objects.

### VIII

So the thesis we have been considering amounts to no more than a restatement of the Principle of Verification. But of course it then amounts to no less than this. Whatever can be said both for and against the one can be said both for and against the other. The only trouble is: The arguments on both sides are excessively familiar.

But it is worth bringing out explicitly that two objections in particular which were made to the Principle of Verification can also be made against Malcolm's restatement of it.

(1) Is this principle really of any use to us? How is it to be decided if an expression *does* satisfy the principle? Suppose a man were to claim to be able to see a new color—for want of a word he calls it "K." He says: "I can't explain to you what it is any more than I can explain to a man who is color-blind what redness is. You just have to be able to see which things are K and which are not. And if you can't, you're just color-blind to K-ness." Well, in fact we don't see any difference between the things he calls "K" and the things he says are not K; the things he calls "K" all look black to us, but many black things, he says, are not K. Is his word "K" a kind-name in a language, or are his utterances in which "K" appears mere noises? The principle tells us "K" is a kind-name in his language only if it is possible to find out whether or not a thing is K. Only if it is possible that *we* should find this out, for *his* memory impressions aren't going to count here. Is it possible that we should find out, i.e., come to be able to see which things are K and which are not? One is inclined to say: It is possible only if "K" is a kind-name in his language, and not a mere noise. So we come full circle.

(2) Connected with this is a second objection. The principle, it will be remembered, was this: A sign "K" is not a kind-name in a man's language

<sup>8</sup> P. F. Strawson (critical notice of Wittgenstein's *Philosophical Investigations*, *Mind*, vol. 53 (1954), January number) explicitly suggests that the notion "private language" may be generalized: i.e., that the terms need not stand for sensations, but may also stand "for things like colours or material objects or animals." But in fact he is concerned with a quite different notion of "private language": on Mr. Strawson's view (cf. pp. 84–85) a private language is one which merely is not, and never has been, understood by anyone but its speaker.

unless it is possible to find out whether or not a thing is K; and let us call this condition on a man's use of a sign "C." What we might then ask is: Is "C" a kind-name in a language? Well, it is a kind-name in a man's language if it is possible to find out whether or not he so uses "C" that it is possible to find out whether or not a sign in a given use does satisfy this condition. Does "K" in the preceding example satisfy it—over and above its perhaps seeming to its user that it does? Do "table" and "chair" satisfy it—over and above its seeming to some non-sceptic that they do? (And this difficulty becomes especially acute if it is strong finding out that is required.) How should I find out whether or not they do—which is not merely to be a matter of my asking myself whether or not it seems to me that they do? And it should be stressed that unless this is possible for some man, then if the new principle is true, "C" is not a kind-name in anyone's language. In which case, what should we take the principle to be saying? One is always inclined to suspect that this kind of question—what happens when we apply the principle to itself?—is something of a game, for can we not learn something important from the principle anyway? All the same, if a thesis is to be defended, then embarrassing difficulties have to be faced.

## IX

In conclusion, I would just like to say this about the new form of the Principle of Verification, that it seems to me very clear that Wittgenstein himself

would never for one moment have subscribed to it. It is astonishing that a man who repeatedly insisted that he put forward no theses in philosophy should constantly be credited with having proved this or that thesis, from the private language thesis to the sorts of things that get called "Wittgenstein's Theory of Meaning," or "Wittgenstein's Theory of Meaning As Use."

There is a vast difference between proposing that we ask a certain question in order to get clearer about the nature of a certain kind of claim, and declaring that if this question does not receive the sort of answer one might have expected, or has no answer at all, then claims of the kind in question are spurious, or meaningless, or not really in the language.

Wittgenstein does say to the Solipsist: You have made a *grammatical* movement, which you interpret as a quasi-scientific discovery. And yet he adds (#401): "But there is an objection to my saying that you have made a 'grammatical' movement. What you have primarily discovered is a new way of looking at things. As if you had invented a new way of painting; or, again, a new metre, or a new kind of song." No doubt hints of this sort are obscure; no doubt it is not clear what it is that we are to do with them. But that is no reason for thinking that we are free to disregard them, or for thinking that they are (perhaps) mere poetic embellishments on what is really a quite straightforward theory about the necessary conditions for meaningfulness or for being-in-a-language.

*Boston University*

### III. RULES AND UTILITARIANISM

B. J. DIGGS

ALTHOUGH moral rules have had a prominent place in recent moral philosophy, their character is not clear. One reason for this is the vagueness and ambiguity which infect the use of the term "rule": Philosophers tend to conceive of moral rules on some particular model, sometimes in a confused way, often innocently and without a clear view of the alternatives. J. Rawls called attention to one important instance of this: He pointed out that the tendency to regard rules as convenient guides, or as summaries of earlier experiences, seems to have blinded some philosophers "... to the significance of the distinction between justifying a practice and justifying a particular action falling under it. . . ."<sup>1</sup>

Partly as a consequence, utilitarianism has been interpreted in a special way, as asserting that the rightness and wrongness of particular acts is decidable on general utilitarian grounds. This form of utilitarianism, so-called "act utilitarianism," is open to serious and well-known objections.<sup>2</sup>

The appeal of the recently more popular "rule utilitarianism" is that it is able to meet some of these objections, and still retain the tie between morality and "the general welfare," which is one of the most attractive characteristics of utilitarianism. I shall argue in this paper, however, that rule utilitarians (and some of their critics, and many others who view moral rules in the same general way) have also tended unwittingly to adopt a particular kind of rule as the model of a moral rule. When this kind of rule has been delineated, and alternatives noted, I think rule utilitarianism loses much of its initial appeal.

My object in this paper, however, is not so much to refute rule utilitarianism as to contribute to the clarification of moral rules. By distinguishing two kinds of rules I shall try to illuminate one of the fundamental options (as well as one of the funda-

mental confusions) open to moral theory. (1) The first kind of rule is exemplified by the rules which workers follow as part of their jobs; these rules may be used to describe a job. (2) The other kind of rule characterizes such common games as baseball, chess, and the like. Both kinds of rule define "practices," but the practices are very different. I think the easy tendency to confuse them may have blinded moral philosophers to significant distinctions between justifying a system of rules designed to contribute to some goal or product, justifying a system of rules which defines a "form of life," and justifying moral rules. Marking these distinctions should help clarify certain steps taken in recent moral philosophy: One should be able to appreciate more fully the point of Baier's assertion that although moral rules are "for the good of everyone alike," they are not designed to promote the greatest good of everyone.<sup>3</sup> One should also be able to see more clearly why Rawls maintains that the decision on the rules of justice is not properly conceived on the utilitarian model, as an administrative decision on how to promote the greatest happiness.<sup>4</sup> The analysis of rules is illuminating, moreover, not only because it helps mark major differences of this kind, but also because it shows what is behind some of the twists and turns of moral theory.

#### I

1.0 The first kind of rule which I shall describe belongs to a large class of rules which I call "instrumental." All rules in this large class are adopted or followed as a means to an end, in order to "accomplish a purpose" or "get a job done." The simplest of these rules is the "practical maxim" which one ordinarily follows at his own pleasure, such as "Be sure the surface to be painted is thoroughly dry" or "Do not plant tomatoes until after the last frost."<sup>5</sup>

<sup>1</sup> "Two Concepts of Rules," *Philosophical Review*, vol. 64 (1955), pp. 29-30.

<sup>2</sup> Cf. e.g., R. B. Brandt, *Ethical Theory* (Englewood Cliffs, N.J., 1959), chap. 15.

<sup>3</sup> K. Baier, *The Moral Point of View* (Ithaca, N.Y., 1958), pp. 200-204.

<sup>4</sup> "Justice as Fairness," *Philosophical Review*, vol. 67 (April, 1958), pp. 164-194. It will be clear that Rawls's analysis in "Two Concepts of Rules" does not support a utilitarian theory.

<sup>5</sup> Cf. Max Black, "Notes on the Meaning of 'Rule'," *Theoria*, vol. 24 (1958), pp. 121-122; reprinted in his *Models and Metaphors* (Ithaca, N.Y., 1962), pp. 95-139.

The instrumental rule to which I call attention is more complex. On many occasions when one wants a job done, either he is not in a position or not able or not willing to do the job himself. If he is in a position of power or authority, or if he has money, he may simply order or hire others to "do the job" and leave it to them. In numerous cases, however, he himself lays down rules of procedure, and establishes "jobs" or "roles" in the institutional sense. A "job" in this latter sense is not a job to be "done," but a job to be "offered to" or "given" to a person. If a person "takes" or is "assigned" "the job," then we often think of him as under an obligation to "do his job," and this partly consists in his following rules. Instrumental rules of this kind, unlike practical maxims, have a social dimension: It *makes sense* to ask whether a job-holder (or role-taker) is *obligated* to follow a particular rule, or whether this is one of his *duties*, and the penalty attaching to a breach of the rules does not consist simply in his not "getting the job done."

Rules of this kind are found in very different institutions. Some are rules of a "job" in the ordinary sense. Others apply to anyone who voluntarily assumes a "role," such as "automobile driver." Others characterize a position which one is obliged to take by law, for example, that of private in the army. The goals which the rules are designed to serve may be ordinary products of labor, such as houses, steel beams, etc.; or fairly specific social goals such as "getting vehicles to their destinations safely and expeditiously"; or goals as general as "the national defense." In some cases the rules, differing from job to job, mark a division of labor, as the rules which say what factory workers, or the members of a platoon, are to do. In other cases, the same rules apply more or less equally to all, as in the case of (at least some) rules regulating traffic.

Notwithstanding their variety, these rules can be

classified together because they share two fundamental characteristics: (1) The rules prescribe action which is thought to contribute to the attainment of a goal. This is the "design" of such rules, at least in the sense that if the prescribed action does not effectively contribute to the attainment of the goal, for the most part, then the rule itself is subject to criticism. (2) The rules are "laid down" or "legislated" or "made the rule" by a party which has power or authority of some kind; one cannot learn "what the rules are" simply by determining what general procedures most effectively promote the goal. This latter characteristic sharply differentiates these rules from what I have called practical maxims, although both share the first characteristic and are "instrumental."<sup>6</sup>

I shall now consider each of these two characteristics in turn.

1.1 Since rules of this kind are designed to serve a goal, the "best" set of rules is that set, *other things equal*, which is most effective in promoting the goal. The qualification is important: One ordinarily asks the question, "Is this a good rule?" in order to determine whether or not the action to be prescribed by the rule, together with other acts, will most efficiently produce the goal, without violating certain other rules, and in a way that harmonizes best with other aims, assuming persons can be persuaded to follow the rule.<sup>7</sup>

Consider a factory planner designing an assembly line, or an army officer considering platoon reorganization, or a traffic planning commission trying to decide whether a street should be made a throughway. In each case rules are proposed, but there is no contradiction in saying that action on the rules will not contribute to the goal. Within its context the question "Is this a good rule?" is one of practical fact and experience. This indicates one sense in saying that the goal is "over and beyond" the action and the rules.

<sup>6</sup> Practical maxims should not be dismissed, however, as "mere rules of thumb" on the one hand, or as "simply stating relations between means and ends" on the other. When one follows a maxim the rule *directs* action and is a *criterion* of certain kinds of rightness and wrongness in acting.

In passing note that Rawls's "summary conception," as a whole, does not properly apply to practical maxims, although several features of this conception do apply. Rawls's analysis, admirable as it is, is very apt to mislead. For the "summary view," as he calls it, is a blend of two quite distinct conceptions: In part it is a confused conception or a misconception of a rule, as a summary or report. In other respects it is an accurate conception of what I have called a practical maxim. This may account for an ambivalence in Rawls's article: Cf. "... it is doubtful that anything to which the summary conception did apply would be called a *rule*." [(p. 23) "Two Concepts . . ."] with "Some rules will fit one conception, some rules the other; and so there are rules of practices (rules in the strict sense), and maxims and 'rules of thumb'." (p. 29). The point is that maxims are rules in a *different* sense from other kinds of rules, whereas no rule, *qua* rule, is a summary or report.

The importance of this point is that there are two possible confusions here, not one: A person may conceive moral rules as summaries or reports, or he may conceive moral rules on the model of maxims. The texts of Austin and Mill, which Rawls cites, together with Rawls's discussion, suggest that the latter, more than the former, was their mistake. V., however, note 13 below.

<sup>7</sup> Cf. my "Technical Ought," *Mind*, vol. 69 (1960), July issue.

There is another sense in saying this: In practice a goal is often described in terms of rules or procedures which are thought to produce it (when, for example, a beam is to be built according to procedural specifications). Moreover, at the time of action one may not be able to say just what he wants in other terms. Nevertheless, there is no contradiction, explicit or implied, in saying that this person got the goal (in the sense that he can truthfully say "This has all the desirable features of what I wanted") without anyone's having laid down or followed rules. Although the beam was not constructed according to specifications, tests may now show that it is as strong as one could have wished for. In this sense it is *logically* possible for one to attain the goal which a set of instrumental rules is designed to serve without these rules having been followed. I shall refer to this characteristic by saying that the goal of any set of instrumental rules is "logically independent" of these rules.

Although an instrumental action is *properly* described in many ways, depending on the context, it can always be *truthfully* described in terms of a goal, as a "trying to get or produce G." For a goal is essential to such action, and to the rules which guide it. Nevertheless, it is clear that it is logically possible to act and follow instrumental rules without attaining the goal, and to attain the goal without following rules.

Moreover, although obviously one cannot act on a rule of any kind if there is no rule, one can act *in the way* specified by a set of instrumental rules (as well as attain a desired result) without *these* rules having been adopted. A group of workers, for example, may hit upon certain procedures which are so effective that they are made "the rule"; in such a case we may say, somewhat misleadingly, that one discovered a good rule by observing the actual results of a line of action. In complex cases it is very unlikely that men will act in the way rules would prescribe if the rules have not in fact been enacted. Nevertheless, there is no contradiction in saying that men acted in this way but there were no rules prescribing this course of action.<sup>8</sup>

Thus in the case of instrumental rules the action as well as the goal may be said to be logically independent of the rules.

1.2 Now consider the second major characteristic of rules of this kind, namely, that they are "laid down," "legislated," "made," or "adopted."

It is clear enough that an employer, for example, who "informs" his employee of the rules, is not

simply "giving information." Moreover, this act or performance is very different from one's "adopting" a practical maxim or making a rule "a rule for himself." Note that in the case of a maxim the adoption of the rule is "incomplete" so long as one simply resolves to follow it. Rules of the present kind, however, are normally made for others to follow: To make their adoption complete, one must get at least some of these others "to agree," in some sense, to follow the rules.

This is so in spite of our sometimes speaking, in the sense indicated earlier, of one's "discovering a good rule" of this kind. We also speak of an administrator's "thinking of a good rule," "deciding on a rule," and "informing an employee of the rules decided on." It is quite clear, however, that "thinking of a rule" and "deciding on it" are steps taken *in the direction of* adopting a rule; the latter corresponds roughly to the stage of "resolution" in the case of a maxim. They are only steps; the rule will not become effective, and strictly speaking, will not *be* a rule, until it is "put in force" or "made a rule."

Legislation is one way of putting such a rule in force. In this case parents and guardians "teach" their children what the laws are; they do not ask for consent. In other cases the members of a group, working co-operatively, "decide on the rules," or an employer or a sergeant "tells one the rules." By such an act those subject to the rules are "directed to follow them," and the rules are then "in force." The rules serve on the one hand as guides to action—they tell one what to do—and on the other as criteria of correctness of action—acts in accord with them are said to be *right* and breaches of them are said to be *wrong*. The rules thus tell one both *what* to do, and *that* he should do it. They are useful just on this account: One may lay down rules of this kind to make use of unskilled labor, or to gain the benefits of a division of labor, or simply to co-ordinate activity as in the case of an efficient traffic system.

The analysis of what the various cases of adopting a rule have in common, and what it is to be subject to rules, takes one to the difficult problem of what constitutes an authority. For our purpose the following will suffice: A party seems to be constituted as a *de facto* authority when one accepts the fact, that this party prescribes an act, as a *reason* for following the prescription (a rule of the present kind being one form of prescription). This indicates the somewhat technical sense of saying that the rule follower

<sup>8</sup> Cf. Rawls, *ibid.*, p. 22.

"agrees to" follow the rules.<sup>9</sup> In the case of rules of the present kind authority is ordinarily constituted, and agreement to follow the rules obtained, by contract, law, convention, or the like. Some such arrangement is necessary to induce a person to follow rules of this kind, since persons other than the rule-follower "are interested in" the goal, and normally he himself does not get (more than a share of) the product of his labor. The contract, law, or convention both promises some reward to the rule-follower, and at the same time converts others' "being interested in" the goal to their "having an interest in it"—in a legal or quasi-legal sense. This, of course, is why one who follows rules of this kind, unlike one who adopts a maxim as his guide, is not free to alter or follow the rules "at his pleasure."

The point which needs particular emphasis here, however, is that the contract, law, or convention is essential to the rule's being a rule; it is not "external" to the rule, since without it one's "laying down the rules" would be only so much rhetoric. When a contract is simply "to do a job," notice that the criterion of correctness is simply "getting the job done." If I hire a person to paint a house, he has done what he is supposed to do when the house is painted. On the other hand, to the extent to which a contract lays down rules specifying how the job is to be done, the rules are the criterion. If a painter contracts to follow certain procedures, and then fails to follow them, he has not done what he is supposed to do. This should make it quite clear that it is the contract, law, or convention which determines in a given case that rules will be the criterion of correctness. The "agreement" secured by contract, law, or convention thus makes a rule a rule, and without something like it there could be no rules of this kind.

1.3 The discussion of the two major characteristics of these rules reveals two criteria of correctness. On the one hand, there is the criterion of a "good" rule. On the other, there are rules *in force* constituting a criterion in certain respects of the *right thing to do*. In the case of these rules there is thus a clear distinction between the justification of a rule or practice and the justification of a particular action falling under it. Perhaps on this very account some have been led to view moral rules as rules of this kind.

1.3.1 Before going on to moral rules let us notice that this distinction is not important simply because acts are judged by rules which are judged in turn in another manner, in this case by reference to a goal. The significance of the distinction derives more from the fact that the two criteria are "independent" in the following way: One may do the thing which most contributes to the goal, yet violate the rules in force; and one may act according to the rule in force when the rule is a poor one.

Moreover, the rules *in force*, not the rules which are *best*, constitute (at least under certain conditions) the criterion of right and wrong acts. This is evident in practice: A worker who does his job is *entitled* to his pay, whether or not the rules he follows in doing his job are *good* rules. This question, whether or not the rules in force are "good," ordinarily does not have to be settled for them to serve as a criterion of right action. Normally it does not even arise.

Of course, one might criticize the rules *in force* as "illegitimate" or as laid down by one who lacks rightful or proper authority, and *on this account* argue that they are not the "true" criterion of right action. However, the question of the "legitimacy" of the rules is not settled by determining which rules are best. To try to have it this way would be to invite disagreement concerning which rules *are* best, and to have no effective rule at all.<sup>10</sup> It would be wholly impractical to accept as authoritative or binding, and as the criterion of right action, only "the rules which are best." Who, for example, would lay down, or contract to follow under penalty, rules characterized only in this way?

Thus, even though rules of the present kind are explicitly designed to promote a goal, the rule follower is not generally at liberty to use the goal as his criterion of the right thing to do. The distinction between the two criteria so far remains firm.

1.3.2 Nevertheless, the independence of these two criteria can be overemphasized. For one thing, the criterion of a good rule, in virtue of its being used by those who adopt rules, is an indirect criterion of right action. The rules which are the criterion of right and wrong action do not prescribe action which just *as a matter of fact* contributes or fails to contribute to the goal; the rules are *criticizable* if they are not good rules. Thus it does not "just so

<sup>9</sup> Cf. Black, pp. 120-121. Black's analysis of the "laying down of rules" in terms of "promulgator activities" and "subject activities" (pp. 139-146) is illuminating, as is H. L. A. Hart's recent analysis of the complex idea of "acceptance" in the case of the law. *V. The Concept of Law* (Oxford, 1961), chaps. IV-VI, esp. pp. 107-114.

<sup>10</sup> Cf. Hume's remarks on the need of a "determinate rule of conduct," or "general rules," in his discussions of justice, both in the *Treatise* and *Inquiry*. Hume, however, does not make precisely the same point.

happen" that the right act *tends* to contribute to the goal. If it did not generally do this it would not be called "right," for there would be no such rules.

Second, no statement of a rule includes reference to all conditions pertinent to its application; one would not wish so to encumber it, even if every contingency could be foreseen. This implies that every rule follower is expected to know "what he is doing" in a sense larger than "following the rules"; and if the rules are instrumental he is often expected to know the goal to which his rule-directed action supposedly contributes—to know "what he is doing" in this sense. Not always, to be sure, but often he could not make a sound judgment of when and how to apply the rule without this knowledge.

For both of these reasons it is a mistake to say, in a pedestrian and casuistical way, that "the criterion of right acts is the rules." It is a mistake to think of *every* exception and *every* case as somehow included in the rule. The motive for doing so, presumably to preserve the authority of rules, is mistaken: There is an important difference between interpreting a rule, or violating it in *special circumstances*, and deciding each individual case just as if there were no rules. A person subject to rules who follows the latter course merits a special kind of criticism. Although it is difficult to specify conditions in which the violation of an instrumental rule is proper, surely the bare fact, "that by doing so one can better promote the goal," is not sufficient. The rule follower is not the sole or final authority on the propriety of breaking a rule, even when it is for the benefit of the other party.

This brings us back to the independence of the two criteria. However, it should now be clear that these criteria are interrelated and operate together. Moreover, since there are two criteria in the case of rules of the present kind, it always *makes sense* to ask if an action right by the rules is also right in the

respect that it is good that a rule prescribes it. It not only *makes sense* to speak of its being proper to violate a rule, "successful violations" tend to be commended.

## II

2.0 As soon as rules of the foregoing kind have been described it is rather obvious that many moral theorists, intentionally or not, have cut moral rules to their pattern. Anyone who regards the standard of morally right action as itself a means to an end will have this tendency, and this is typically true of rule utilitarians: The distinctive characteristic of their theory is that a system of rules is the criterion of morally right action, and these rules in turn are to be judged good or bad according to the consequences which action on the rules either generally produces as a matter of fact, or would produce if people could be persuaded to follow them.<sup>11</sup> The consequence which has been thought to be critical in assessing the soundness of a system of rules has been variously identified, as "the happiness of all," "public utility," "security," "the general welfare," etc. Nevertheless, in spite of the difference in name and even in conception, this has been taken to be a consequence, real or possible, and as an end or goal which a good system of rules would first promote and then ensure. The question of which system of rules will be most successful in this respect generally has been thought to be, at least broadly speaking, empirical: Fact and practical experience will decide which system is best. The theory thus implies that the goal, and goal promoting action, both, in senses indicated earlier, are *logically* independent of any system of rules. This fundamentally instrumental and telic character of the system of rules, and indirectly of rule-directed action as well, is a distinctive feature of utilitarianism.<sup>12</sup> Moreover, as

<sup>11</sup> See, for example, J. O. Urmson's "The Interpretation of the Moral Philosophy of J. S. Mill," *Philosophical Quarterly*, vol. 3 (1953), pp. 33-39. By and large I agree with this interpretation of Mill, although Mill showed other tendencies, not only toward a more radical utilitarianism but, in the opposite direction, toward the ethics of Bradley. John Austin is sometimes said to be a good representative of this point of view, but his conception of moral rules as commands, learned in the way we learn practical maxims, is a hodgepodge (see *The Province of Jurisprudence Determined*, Lectures I-III). In some respects Hume's discussion of the artificial virtues, especially justice, is a much better (and perhaps the best) classical example of this type of theory.

Among contemporaries (and apart from useful textbook presentations: see Brandt, loc. cit., and J. Hospers, *Human Conduct*) S. Toulmin in *The Place of Reason in Ethics* and P. H. Nowell-Smith in *Ethics* have come closest to an explicit statement of the theory.

An examination of actual cases of this kind of theory, with all the proper qualifications, especially if the theory is extended beyond utilitarianism, would require considerable space. I do not undertake the historical investigation here. In my judgment, the theory has a popularity which exceeds its merit, and some tendencies which are pernicious (see Section IV below). By isolating the germ, the disease may be better understood—its valuable antibodies notwithstanding.

<sup>12</sup> It would be a mistake to say that utilitarians maintained this deliberately, after considering alternatives, or even that they did so consistently. John Stuart Mill, in Chapter IV of *Utilitarianism*, seems to have been unaware of the issue when he discussed happiness as "a concrete whole" and virtue as one of its "parts." Cf. below 4-5.

I pointed out above, it is an essential feature of rules of the foregoing kind that persons other than the rule follower are "interested" in the product; this "interest" is expressed in some kind of contract, convention, or law which gives the rules authority. In utilitarian theory the "party-in-authority" tends to be "the people"; directly or indirectly they enter conventions, "adopt" rules, then enforce them, so that all may share the fruits of the rule-directed action. The product is shared, the goal is the good of all.

2.1 Moral rules on the rule utilitarian view thus have the basic characteristics of the rules which I discussed in (1). When the two are compared, and the analysis in (1) is brought to bear, it quickly reveals that rule utilitarianism is faced with a fundamental problem. If the position is to have the advantage over act utilitarianism that is claimed for it, then the criterion of right action must be a system of rules and not general utility. Rules are a criterion of right action, however, only on condition that they are "rules-in-force" and in some sense "agreed to." But obviously the rules which are "in force" or "agreed to" may or may not be the rules which maximize utility; and to the extent that they are not, then the "best rules" by the utilitarian standard, not having been "adopted," are not the criterion of right action. The best rules may not even be known. The "rules" and the "utilitarianism" in "rule utilitarianism" thus constitute two independent criteria, and they may not be in much accord.

2.1.1 The analysis in (1) not only clearly shows the nature of this difficulty, but also helps one to understand some of the directions in which utilitarianism has moved in an effort to avoid it. Some good utilitarians, mindful of evil in ordinary conventions, tend to say that just as men *ought* to adopt a rule only if it maximizes utility, so one is *obligated* to follow a rule only if it maximizes utility. This doctrine implies that one may freely disregard a rule if ever he discovers that action on the rule is not maximally felicitous, and in this respect makes

moral rules like "practical maxims." It deprives social and moral rules of their authority and naturally is in sharp conflict with practice. On this alternative rule utilitarianism collapses into act utilitarianism.<sup>13</sup>

2.1.2 Other rule utilitarians, equally concerned to avoid an ethical conventionalism, either close their eyes to the difficulty or else overlook it. They either just declare an ideal set of rules to be the criterion, or else say that the criterion of right action is the system of rules which, *if* adopted, *would* maximize utility, or something of the sort. Such a formulation clearly does not acknowledge that rules must be adopted if they are to be rules: The "if adopted" is only a way of describing the ideal and actually obscures the necessity of a rule's being adopted.

The fact that it is commonly the case that some moral principles and rules to which a person subscribes are not "in force" in his society raises important issues for *any* moral philosophy of rules. I cannot even try to do them justice here. Nevertheless, surely it is a mistake to maintain that a set of rules, thought to be ideally utilitarian or felicitous, is the criterion of right action. If the rules are simply described in this way, and are not enumerated, we so far do not have any rules and are not likely to get any.<sup>14</sup> On the other hand, if we are presented with a list, but these are not rules in practice, the most one could reasonably do is to try to get them adopted. A manager in the quiet of his office may dream of a system of rules which will maximize production, and a utilitarian may build a theory around the set of rules which will maximize utility. Surely the latter would be as foolish as the former if he said that these ideal rules are the criterion of right and wrong acts. As previous analysis has shown, acts are not judged by proposed rules, ideal rules, and rules-in-theory: for these do not fully qualify as rules.<sup>15</sup>

2.1.3 Other rule utilitarians show a finer appreciation of the logic of their position: They interpret moral rules on analogy with the rules in (1), even

<sup>13</sup> For a clear recent statement of this position, see J. J. C. Smart, "Extreme and Restricted Utilitarianism," *Philosophical Quarterly*, vol. 6 (1956), pp. 344-354. Notice that Smart argues explicitly that moral rules are "rules of thumb."

<sup>14</sup> Cf. above, 1.3.1.

<sup>15</sup> See 1.2 and 1.3.1 above. Since utilitarianism is rather often associated with reform, it tends to be formulated in ideal terms. See, for example, J. S. Mill's most explicit statement of his position in Ch. II, paragraph 10 of *Utilitarianism*: "... the standard of morality, which may accordingly be defined 'the rules and precepts for human conduct', by the observance of which an existence such as has been described might be, to the greatest extent possible, secured to all mankind..." In this passage, how is "possible" to be taken? Does it mean "possible, within the framework of existing institutions?" For one attempt to avoid in this way the difficulties inherent in an ideal formulation, see R. B. Brandt, *op. cit.*, pp. 396-400. This attempt goes only part of the way in meeting the difficulty. On the difficulty itself cf. H. J. McCloskey, "An Examination of Restricted Utilitarianism," *Philosophical Review*, vol. 66 (1957), esp. pp. 475-481; and J. Austin, *op. cit.*, Lecture III.

if it forces them to admit that the criterion of right action is not the set of rules which maximizes utility. This alternative seems to be popular with those whose primary allegiance is to a "morality of rules," and who are utilitarian only because they suppose that "welfare" *must* have something to do with morality. (After all, what else *can* serve as a criterion of rules?)

On this alternative it always makes sense to ask whether or not a "moral or social convention" subscribed to in practice is best, and this gives sense to the question, sometimes asked, whether a people who follow their conventions act in the best way they could. At the same time the question, whether an individual ought to do something in particular—for example, repay money borrowed—is quite a different question, to be answered by referring, at least in part, to the practices and conventions of that society. Such a view does not make the blunder of taking an ideal system of rules as the criterion of which particular acts are right, and yet it does not endorse conventions which are obviously questionable. One may seek earnestly to reform the moral conventions of a people, and yet insist that these conventions, some of which are in need of reform, are the general criterion by which a man must decide what in particular he ought to do, and by which his acts are to be judged. At the same time, such a view need not dichotomize the two criteria. As we found above, rules of this kind have an open texture which permits the criterion of the rules to enter into their proper interpretation. I think we may presume, moreover, that there are instances in which one should violate the letter of a moral rule when following it would clearly be to the detriment of the general welfare, or the welfare of all parties concerned. Rule utilitarians could no doubt take instances of this sort to support their theory. As we also found above, one may admit this without depriving rules of their authority.<sup>16</sup>

### III

3.0 A careful development and criticism of rule utilitarianism, as just outlined, would be worth while, but it is outside the range of this paper. Even without this development, however, it can be shown that rule utilitarians, by using the kind of rule in (1) as a model, have exercised a definite option, and I want to indicate the general character of this option. To do this, I shall first consider briefly the rules of certain kinds of games.<sup>17</sup>

3.1 Rules of common competitive games, such as baseball, chess, and the like, say how a game is to be played. They state the "object of the game," "the moves," "how the counting should go," etc. Often they are stated in "rule books," and sometimes they are enforced by referees appointed by an acknowledged authority. These formalities, however, are not at all necessary. The rules must be "laid down" or "adopted" in some sense, but all that is required (in the case of those games being discussed) is that a group of players "agree" on a set of rules. This agreement may consist simply in their following and enforcing rules which they all have learned: Think, for example, of a group of small boys playing baseball, and think of the difference between one's knowing the rules and playing the game. In such cases there is no formally agreed-upon authority; each player—in principle—is both rule-follower and rule-enforcer. No player has the authority to modify the rules at will, but the players together can change them in any way they see fit. As one should expect, there are many variations.

In the latter respects game rules of this kind are quite like the rules in (1). These game rules, however, noticeably lack the first major characteristic of those rules: They are not designed to yield a product. More precisely, they are not adopted to promote the attainment of a goal which, in the

<sup>16</sup> I think this is the most favorable interpretation which can be given to the utilitarianism of the nineteenth century reformers: They framed a theory which would make sense of reform, but at the same time had too much practical (if not always philosophical) sense to advocate the use of the criterion of rules as the criterion of acts. It is as if they perceived the importance of moral rules and practices but were unable fully to accommodate these to their theory. I think that the presence of the two criteria, which the analysis of the rules in (1) clearly reveals, explains for example the "tension" between chapter two of Mill's *Utilitarianism* on the one hand, and chapters three and five on the other.

<sup>17</sup> I can be brief because rules of this kind have been discussed by others. I shall mostly confine myself to points not previously mentioned, or at least not emphasized. I am perhaps most indebted to Rawls's acute analysis of what he calls the "practice conception," and on the whole agree with it. The name is misleading since very many "practices," as we ordinarily think of them, are defined by rules (e.g. by job rules) which are quite unlike those to which his "practice conception of rules" properly applies. Although unimportant in itself, it is just this kind of thing, I suspect, which has led moral philosophers into serious error. One can sympathize since it is almost impossible to find a conventional expression which is not misleading in some important respect.

senses indicated earlier, is "over and beyond" the rules.<sup>18</sup> They do not serve a goal which is "logically independent" of the game which they define.

3.1.1 Of course people who play games do so with various motives, and some of the goals which motivate them are logically independent of the game; for example, exercise, recreation, the opportunity to talk to friends or make a conquest. Undoubtedly games are popular because they serve so many ends. Nevertheless, motives and goals of this kind are not essential. Many players participate (so far as can be determined without psychoanalyzing them) "just because they want to" or simply "from love of the game." Actually this kind of motive, even if it is not typical, is that which is most distinctive of players: One who "loves a game" commonly regards another, who lacks the motive, as poorly appreciating "the quality of the game." This is apt to be missed just because games have been turned into instruments, for exercise, diversion, etc., to such a great degree. The point is, they *need* not be.

Moreover, games *qua* games do not seem to have a design or goal *different* from the motives of the rule-followers, in the way rules of jobs commonly do. What is this goal? One who most appreciates a game speaks about it rather as if it were an aesthetic object, worth playing on its own account and apart from any product or result; and if he is asked to justify his claim that it is good, he seems to have a problem analogous to that of justifying an aesthetic judgment.<sup>19</sup> Sometimes, to be sure, the rules of games are changed, and in particular instances violated, in order to change the consequences. Many official rules, for example, have been changed in order to lessen player injuries; and particular persons may find a game played by the official rules too strenuous, or pursuit of the ball after a bad drive too troublesome. These facts, however, do not imply that the rules are designed to produce consequences, such as the right amount of exercise or exertion, or the good health of the players. Changes of the kind mentioned simply indicate that the rules of a game, like the rules of a job, are adopted in a context by persons who have many desires and many obligations other than "to play the game" and "follow its rules." Games are often altered to make them harmonize better with

such contextual features. It is true, of course, that persons who have turned games into instruments change or violate the rules more readily. As we say, these people do not take the game as seriously.

Some philosophers are inclined to say that even when one plays a game "just because he wants to" or "for love of the game," the game is still an instrument—to "his enjoyment" or "pleasure." This stand depends for its cogency on our being able to describe this pleasure or enjoyment without referring to the game, which should be possible if the pleasure or enjoyment really were something separate from playing the game. However, although it is clearly possible to play a game and not enjoy it, the converse does not appear plausible. To be sure, one sometimes says that he gets about the same enjoyment from one game as another, especially when the two are similar. But this is apt to mean that he has no strong preference for one game over another, that he likes one as well as the other, not that there is a kind of pleasurable feeling which in fact results from both, more or less equally, and which *conceivably* could be had from very different activities or even from being acted *on* in some way. (Similarly, when one says that he "likes to talk to one person about as much as another," this clearly does not mean that talking to the two persons produces the same kind of pleasure in him.) Moreover, when we speak of getting about the same enjoyment from two games, sometimes the "enjoyment" does not appear to be, strictly speaking, the enjoyment "of playing the game," but rather the enjoyment of exercising, talking to friends, etc. I do not deny, however, that games can become instruments. I want to argue that they need not be, often are not, and that in calling them games we do not imply that they are instruments.

The kind of goal the pursuit of which to some degree is essential to the playing of the game is the "object of the game," as defined by the rules, and the various sub-goals which promote this object according to the rules. Such goals as these, for example, "to score the most runs," "to get the batter out at second base," obviously are not logically independent of the rules of the game—if there were no rules it would be logically impossible to try to do these things. It is just nonsense to speak of

<sup>18</sup> Some games have become instruments to such a considerable degree, and some instrumental activities have become so much like games, that no description will prevent the intrusion of dubious and borderline cases.

<sup>19</sup> This reminds one of the ancient distinctions between "doing" and "making," and between (what the medievals called) "immanent" and "transitive" activity. I do not mean to deny that some jobs are worth doing "on their own account," but even when "one enjoys a job," there is a discernible purpose which it is designed to promote.

changing the rules so that one can better attain the object of the game.

3.1.2 Since the action within a game is designed to attain goals defined by the rules, the action as well as the goal logically depends on the rules: In important respects a move in the game has the consequences it has because the rules say it has; *in these respects* the rules define the consequences and determine the character of the action.<sup>20</sup> Since the character of instrumental action is fixed at least partly by the goal which the action is designed to serve, the action can be described in this essential respect, as a "trying to get the goal," without referring to or presupposing rules. In the case of play in a game, unless the game has become an instrument, this is not possible; if one describes the action in a game apart from the rules, as a "trying to catch a ball," he leaves out the design. On account of this difference one may feel inclined to say that whereas rules of the kind described in (1) *may* be used to describe an action, game rules by defining new kinds of action just constitute "forms of life."<sup>21</sup>

3.2 However, this is but one side of the story, and if it were the only one it is not likely that the two kinds of rules would be confused. To see the other side, which is equally important, one should attend to the fact that the play in a game is not wholly defined by the rules of the game. "The kind of game he plays" ordinarily does not refer to the game as defined by the rules; "to play a game" ordinarily means more than following the rules. The point is that although the object of the game is defined by the rules, since the action in a game normally consists in "trying to attain that object," and since the game rules do not determine success in this respect, the action in *this* respect is instrumental. Players often develop tactics and strategies and skills in playing. Sometimes they follow what I have called practical maxims, and at other times they follow team rules agreed on among themselves or laid down by the "manager." The latter are, of course, examples of the rules described in (1). Obviously they should not be confused with rules

of games, as I have described them. For one can be said to play a game without his following any particular set of instrumental rules.

The point of greatest importance here is that although game rules are not themselves instruments, they support, as it were, a considerable amount of instrumental activity, much of which logically could not be carried on without them. To play a game is typically to follow the rules of the game *and* engage in this instrumental activity; a "good player" does more than just follow the rules. Even one who "loves the game for its own sake" derives his satisfaction from the kind of *instrumental* activity which the rules of the game make possible. Games make new goals, new pursuits, and new skills available to men.

In this situation it is not surprising that some should regard games themselves as instruments. To regard them in this way, however, would be to confuse their function.

#### IV

4.0 The rules of games just considered differ most significantly from the rules described in (1) because they are, by our criterion, "non-instrumental." This point of difference between the two kinds of rules is one of the most important to be found. I have been concerned to mark it here to focus attention on the thesis, maintained by many utilitarians, that moral rules and social institutions are instruments designed to promote a goal logically independent of the rules and institutions. The thesis is only rarely discussed, and I think that failure to discuss it helps account for the recurrent popularity of utilitarianism. However, morality is obviously not a game, and if the thesis is to be fully assessed, moral rules must be carefully analyzed and alternatives considered. This is out of the question here. In the remainder of this paper I shall note a complexity which is too often overlooked, and just indicate the critical force of certain recently developed lines of argument. However, the fundamental issue here is not at all new.<sup>22</sup>

<sup>20</sup> This is the point which Rawls emphasized.

<sup>21</sup> Cf. A. I. Melden, "Action," *Philosophical Review*, vol. 65 (1956), pp. 523-541.

<sup>22</sup> Historically one perhaps first senses the issue in his reading of Plato and Aristotle. Is man's end somehow "writ in his nature" in such a way that it can be determined apart from a determination of virtue? If so, it might be reasonable to regard virtue as a *means* to the end, and instruction in virtue as a matter of learning from practical experience the best means. On the other hand, if man's end cannot be determined without the determination of virtue—if man's end is properly defined in terms of virtue, as activity in accordance with it, and man's nature is defined as potentialities for this end—then virtue is not a means and its discovery in practical experience must be understood differently. Although the second interpretation is the sounder, there were tendencies in medieval thought to favor the first—undoubtedly deriving from the fact that God, who is certainly different from man, was said to be man's end. Moreover, the desire of God was said to be implanted in man's nature. This inclination was said to be a natural participation of the eternal law, and natural virtue was said to be an insufficient means to

4.1 Consider the rule "Do not cheat." Often it is taught in the context of a game, and it acquires a rather specific sense in this context. The rule in this use can be paraphrased as "Do not violate the rules of the game in order to gain an advantage for yourself." In this use the rule logically presupposes games as social institutions; if there were no games, the rule could not have this use and this meaning.

The same general point applies to many other moral rules, such as "Keep your promises," "Do not steal," and "Do not lie." Each of these logically presupposes institutions and practices, such as "promising," "a system of property," "a language." Since these moral rules presuppose such practices, they cannot be understood apart from them; the practice, constituted by its own rules, makes the moral rule meaningful. Philosophical analyses which have attempted to clarify moral rules apart from institutionalized practices have surrounded them with theoretical perplexities and turned them into "mere forms" of morality.<sup>23</sup>

However, the fact that these moral rules presuppose institutions or practices does not *in itself* decide the question whether or not they are instrumental and utilitarian. In some respects the rules "Do not cheat," "Do not lie," etc., are like the rules "Do not violate traffic lights," "Do not drive on the wrong side," etc. These rules obviously presuppose practices, and the rules and practices appear to be primarily instrumental and utilitarian. We can easily conceive of the practices being changed in order to provide a more effective system of traffic control.

On the utilitarian view moral rules and the institutions which they presuppose are rather like a system of this kind. The assumption is that men have various destinations which they want to reach and the social aim is to provide the system of institutions which will be most effective in helping them along. As men together devise such public instru-

ments as roads and bridges, which no one alone could construct, and then regulate the use of these instruments for the "public good," so on this view men together have developed such institutions as "promising," "a system of property," etc. These institutions may not have arisen through deliberate design, although (there often seems to be the assumption that) if an institution or practice has arisen, then it *must* have been rewarding, and consequently *must* have served some purpose. The instrumental character of these institutions is evidenced more directly, however, by the fact that persons hold and dispose of property, make promises, and, quite generally, engage in the life of their institutions with goals in mind. If these reasons are decisive, moreover, one's language, too, should be viewed as a social tool.<sup>24</sup> Certainly men have purposes in speaking.

As in the case of a traffic system, however, on occasion it is to a person's advantage to break the rules of their institutions. Men must be taught not to; they must be made to realize that temporary advantage is far outweighed by the more permanent benefits to be gained if all can be depended on to follow the rules. Moral rules, such as "Keep your promises," "Do not steal," "Do not lie," like the rules "Always obey traffic signals," "Do not drive on the wrong side," seem to be conceived as deriving from the occasional but recurrent conflict between private advantage and public institutions. Utilitarians commonly make the point that if a person in his own interest is sometimes led to violate a rule, he will nevertheless insist, also in his own interest, that others follow the rule: The "security" which derives from a system of public institutions is given an important place in moral theory. Moral rules of this kind thus seem to be conceived as supports for and ancillary to the public institutions which they presuppose. If these rules could only be made to serve a system of truly

God. I think myself, however, that the second interpretation gives a sounder account of the ethics not only of Augustine but also of Aquinas. Yet it is not surprising that out of this tradition there should have come the contrary (Lockian) doctrine that natural law applies to man in a "state of nature," and that men by compact make societies as a remedy for natural evils and as a means to natural goals. This doctrine in turn, by way of reaction, stimulated theories according to which the distinction of right and wrong is not founded in nature, but in contract, convention, or rules. In the nineteenth century the opposition between the two general points of view assumed more of its original form when idealists worked out their own interpretation of the social contract, and opposed utilitarianism. (See, for example, Bradley's "Pleasure for Pleasure's Sake" in *Ethical Studies* and Bosanquet's *Philosophical Theory of the State*.) Very recent philosophy in some respects strongly resembles idealism, undoubtedly because it itself is a reaction to a kind of philosophy which arose in reaction to idealism. For one example, cf. Bosanquet, op. cit., with A. I. Melden, *Rights and Right Conduct* (Oxford, 1959).

This is, of course, only a fragmentary account of the historical origins of the issue.

<sup>23</sup> This misinterpretation accounts for some criticisms of a morality of rules. Cf. A. Macbeath, *Experiments in Living* (London, 1952), Lecture XIII.

<sup>24</sup> Cf. Hume's *Treatise*, III, II, II. Esp. p. 490 in Selby-Bigge edition.

rational (i.e., utilitarian) institutions, the aforementioned conflict would be minimized, as the happiness of all was promoted. The negative morality of rules would be lost in liberal affection for the general welfare.

4.2 Moral rules of this kind in a sense do *tend* to support the institutions and practices which they presuppose: They *tend* to receive their effective interpretation from the character of the institutions, and they are both taught and reaffirmed most vigorously when persons from self-interest show an inclination to violate the rules of the institutions. As a consequence (and for an additional reason which will soon be apparent<sup>25</sup>) these institutions and practices have, as it were, a "moral dimension" or a "moral part." Nevertheless, in assessing rule utilitarianism it is important to distinguish moral rules on the one hand from other rules which also define and characterize the underlying institutions and practices. For it is possible to learn the rules of a game, and to play the game, without being tempted to cheat, without grasping the concept of "cheating," and without learning the moral rule "Do not cheat." It is not uncommon for children to do this. Children ordinarily also learn to speak correctly, in the sense of learning many rules of the language, without learning the rule "Do not lie," thus without grasping the moral concept of a lie. It may not be so evident, but it is also the case that one can learn many rules governing property, can learn to make a promise, etc., without grasping the moral force of the rules "Do not steal," "Keep your promises," etc. There are surely legal experts on property and contract who have, as we say, very little moral understanding.<sup>26</sup>

In considering the soundness of rule utilitarianism, there are thus two interrelated questions. The first is whether or not the institutions of promising, property, language, etc., are instruments serving goals logically independent of these institutions. This bears on the question of the soundness of utilitarianism not only as a *moral* but as a *social* theory. Then there is the more restricted question

whether rule utilitarianism offers a sound account of moral rules.

4.3.1 Several lines of thought, some recently developed, bear on these questions. To take one example, primarily as it applies to the first of the questions: Utilitarians, as already indicated, have put considerable emphasis on "security," if not as *the* goal, nevertheless as an important "part" of the goal. A person cannot be "secure," however, without being able to *count on* others to act and refrain from acting in a variety of ways. His counting on others, moreover, is in a great many cases not "an expectation" based on an ordinary induction. For most often the expectation involved in one's counting on another is based on the fact that the action or restraint in question is governed by rules which define rights, obligations, duties, etc.: One can count on another because the other (presumably) is acting on such rules.<sup>27</sup> For this reason the expression "counting on another" in many occasions of its use makes no more sense apart from rules than "deciding to act" or "acting" makes apart from reasons for acting. There is also the related point that the action which one counts on another to do, itself, in many cases presupposes rules; for example, just as one could not count on a person to "play first base" if there were no game of baseball, so one could not count on another to "keep his promise" or "respect property" if there were no practice of promising or institution of property.<sup>28</sup> Although "security" is an ambiguous term, in the sense in which it refers to a significant social goal it could not mean what it does without rules which define institutions and practices.

For both these reasons "security" just does not appear to be a goal which is logically independent of the rules of institutions and practices like property, promising, language, etc. Moreover, it would seem very strange to think of the greatest number having the greatest happiness or pleasure or welfare without being fairly secure. The utilitarian position thus appears to be quite vulnerable, even apart from the fact that its proponents have

<sup>25</sup> See 4.5 below.

<sup>26</sup> Although an adequate description of property and promising in a sense implies that theft and promise-breaking are morally wrong, a person may fail to "see" the implication. When we teach a child what property and promising are, we commonly say that it is wrong for him to take what belongs to another and wrong for him not to do what he has promised to do. So far, however, the child is not guilty of theft or promise-breaking, and until he has witnessed them, or an inclination thereto, in himself or another (since he has not yet had occasion to *use* the rules "Do not steal" and "Keep your promises"), he will have little practical understanding of these rules. Before he reaches this point, however, he may have learned enough of the underlying rules to exchange property, make promises, etc. Growth in moral understanding is long and complex and participation in ordinary practices does not wait upon it.

<sup>27</sup> Cf. Hart, *op. cit.*, pp. 54-7.

<sup>28</sup> Cf. Hume, *loc. cit.* Black and many others make the same point.

notoriously failed to give "happiness," "pleasure," "welfare," and the like the clarity of meaning which they must have to function as goals.

4.3.2 Furthermore, as the earlier analysis of games revealed, the fact that one does many things as a means to an end when engaging in a practice gives no support to the claim that the practice itself is a means. The fact that one uses various devices to win a game does not imply that the game is an instrument, and similarly, the fact that one uses words as tools, or makes a promise or deals in property for some purpose, does not support the view that institutions and practices such as language, promising, and property are instruments for the promotion of goals logically independent of these institutions and practices. Nor does this appear plausible: It seems rather to be the case that institutions and practices create or establish most of the goals which men pursue, in the sense that these goals, like the object of a game, would be logically impossible without the institutions and practices. It also appears that persons who engage in business, or make speeches, or follow intellectual pursuits ultimately because "they just enjoy doing these things" are rather like players who enjoy a game for its own sake—in the respect that they derive their enjoyment from instrumental activity which is also made possible by institutions and practices.

At this point, however, it becomes apparent that much requires to be worked out before one can replace the utilitarian view of social institutions with another which is more adequate.

4.4 When one turns to consider utilitarianism as a theory of moral rules, to some extent the same arguments apply. For some moral rules are in some respects ancillary to the practices and institutions which they presuppose, and in so far as this is the case, then generally speaking moral rules are just as utilitarian as, and no more utilitarian than, these practices and institutions. Notice that the most common uses of the moral rules "Do not lie," "Do not steal," and the like presuppose not only underlying institutions and practices, but also, as suggested above, a tendency or inclination of some persons at some times not to conform to the institutions and practices. This seems to explain why persons living in a law-abiding community use these moral rules so little. This in turn suggests that moral rules are "protective devices," rather like a police system, which also is little used in a law-abiding community and which also presupposes both institutions and an inclination on the part of some persons to violate them. The "police" view of

moral rules is partial, but it is also partly true: It helps one see why moral rules are so often conceived as "external" to an individual, imposing restraints on him (and why some philosophers tend to pattern moral rules on rules in a prison!) At the same time it helps one understand why some people "internalize" moral rules in the way they do. For some insist on the importance of following moral rules only because they value a system of institutions and the "happiness and security" which the institutions afford. Seeing that valued institutions would cease to exist if people generally did not act in the way moral rules prescribe, they teach these rules—although morality for them is primarily a matter of promoting individual or public welfare, and it would be better if moral rules had little use. This interest in morality is epitomized in the person who regards moral rules as a protector of life, liberty, and property; breaking the rules breeds fear, ruins business, and disrupts the game. This is the internalization of moral rules as ancillary to institutions; it tends to characterize utilitarians past and present.

4.5 Moral rules, however, may be internalized in quite another way, and on this account utilitarianism as a moral theory is open to an additional criticism specific to itself.

For a person who values an institution constituted by rules may come to see that rules by nature apply to all members of a class. One who sees this may then be led to look upon the rules which characterize some particular institutions and practices not simply as "applying to all," but at the same time as constituting "a common standard of correctness." And in this way one may be led to the abstract but practical conception of "a community of men living under the idea of law," of which particular institutions afford so many possible examples. In so far as one thinks that others as well as himself act under this conception, he will no doubt value a particular game or language or any other such institution not only *qua* game, *qua* language, etc., but also as a particular instance and a particular form of such a community.

When the idea of such a community is attained and made to govern practice (as it seems to have been, for example, by the Socrates of the *Crito*) then the moral rules "Do not lie," "Do not steal," etc., will appear in a new light. One who acts under such an idea will teach these rules neither as primarily negative and restraining, nor primarily as supports or protections for particular institutions. For although he may view the rules in these

ways, he will regard them primarily as affirming in so many different ways the fundamental principle "Live under the idea of law." The principle may be stated negatively, in the form "Do not make an exception of oneself," but his primary aim in teaching the rules will be to raise one to the conception of a moral community. Since such a community potentially includes all men, part of the challenge may be to find particular institutions in which the conception can be realized.

Moral rules regarded in this way of course still presuppose particular institutions and practices. However, they are no longer, properly speaking, "ancillary to" the institutions and practices: They now "add something" to the institutions and practices which they presuppose; the institutions and practices now have a new dimension. Cheating comes to be deplored not primarily because it tends to disrupt a game but because it detracts from the quality which a game can have. If there is cheating, one may simply prefer not to play. In a similar way, lying may be deplored because it detracts from the quality of speech, theft because it detracts from the quality of exchange, etc. Put affirmatively, the idea of a moral community is realizable analogically—only in a variety of forms—in sportsmanship, morally mature speech, honest argument, etc. It should be evident that common institutions and practices are often not in fact logically independent of morality; one has to form a limited or abstract conception of them to make them so.

When moral rules are regarded in this way,<sup>29</sup> then obviously they do not serve a goal logically independent of themselves. In the language of Mill, virtue has now become a "part" of the end, a "part of happiness." Only it is clear that when Mill said this, with his usual willingness to sacrifice theory to good sense, he deserted utilitarianism.

The instrumental and utilitarian pattern just will not fit.

## V

Further discussion of moral rules is beyond the aim of this paper. My primary purpose has been to contribute to the clarification of moral rules by clarifying a fundamental option open to moral theory. To this end I have both analyzed the general utilitarian view of social rules and practices, along with some variations, and I have tried to lay bare the (largely implicit) utilitarian view of moral rules. I have analyzed moral rules, however, only to the point where the character and significance of the option, and the force of some of the arguments which apply, will be fairly clear. I do not want to suggest that all moral rules are like those which I have considered. The analysis of games, in distinguishing the moral player from the good player, may remind one that there are two traditions in the history of ethics, one emphasizing an exoteric ethic and a moral law known to all, the other an esoteric ethic and a virtue reserved for the wise. I have been concerned, almost exclusively, with the former, and not all of that.

In the course of the discussion attention has been called to the fact that moral rules can be (and thus tend to be) conceived as summaries, reports, practical maxims, rules designed to promote a goal, rules which define institutions, rules which protect institutions, and as particular forms of the fundamental principle of justice.<sup>30</sup> Marking the important differences between these alternatives should remove more than one confusion and at the same time provide *some* of the subtlety which will be needed if the discussion of moral rules is to make genuine advances in the future.

<sup>29</sup> Cf. K. Baier, *op. cit.*, pp. 200–204, and W. D. Falk's comments on "natural obligation" and "mature moral thinking" in "Morality and Convention," *Journal of Philosophy*, vol. 57 (1960), pp. 675–685.

<sup>30</sup> The list is not meant to be exhaustive. Cf. e.g., D. S. Shwayder, "Moral Rules and Moral Maxims," *Ethics*, vol. 67 (1957), pp. 269–285.

## IV. EMOTION AND THOUGHT

IRVING THALBERG

MANY emotions appear to be founded or based upon states of belief. If we hear that a cattleman is dismayed, anxious, or indignant about falling meat prices, have we any need to inquire whether the cattleman has thoughts concerning the price of beef? It seems perfectly obvious that he must not only imagine that the market value of beef is decreasing; he must be relatively sure of it, if that is what alarms, disquiets, or shocks him. Although it is so obvious that such emotions are founded upon some form of thought, it is not immediately clear what kind of dependency this is. And that is one reason I wish to examine the relation between a person's feelings, moods, inclinations or attitudes, and his convictions, doubts, or guesses. A more important reason for investigating this liaison between emotions and thought is that our pre-analytic view of it will neither accommodate nor show the impossibility of some baffling cases. Let me convey my perplexity by asking whether any of the following statements could be true of John, a dinner guest:

- (a) John is embarrassed that he is late for dinner, but he doubts that he is (late for dinner);
- (b) John is delighted that there will be champagne with dessert, but he merely conjectures that there will be;
- (c) John resents the hostess for having gossiped about him, but he is not at all sure that she has done so.

There seems to be a discord, if not an inconsistency, between John's feelings and opinions. Is it possible to feel embarrassment over something which you believe not to have happened? Why rejoice about a future event when you have no assurance it will occur? Does it make sense to bear a grudge against a person when you are uncertain that the person has wronged you? One reply, which only deepens our puzzlement, is that we have no difficulty at all with some descriptions of closely analogous states of mind:

- (a') John is worried that he is late for dinner, but he doubts that he is;
- (b') John hopes that there will be champagne with dessert, but he merely conjectures that there will be;
- (c') John suspects the hostess of having gossiped about him, but he is not at all sure that she has done so.

If (a')-(c') describe possible *mêlages* of feeling and cognition, why are we inclined to reject (a)-(c) as patently ridiculous? In general terms, the problem I wish to raise concerning the relation between thought and emotion is this: Must we think something whenever we have an emotion? And if so, must each affective state be founded or based upon a particular type of thought? The architectural metaphor has no importance in my analysis, but I do find it less tendentious than some alternative phraseologies: "Must one's feelings be consistent with, ruled or governed by, or in harmony with one's surmises, assumptions, convictions, and doubts?" These logical, political, and musical idioms are quite as inadequate as the architectural one to reveal any blatant or hidden nonsense in the examples (a)-(c). For if apprehensiveness is consistent with (ruled or governed by, in harmony with, built upon) doubt, why does embarrassment clash with disbelief? Similar questions may be asked regarding the other pairs of cases.

These questions, as well as the standard replies to them, are peculiar. This suggests that we must begin with a brief inquiry into the concept of emotion. Then it will become clearer how, and why each sort of emotion—shame, anxiety, and so on—requires a particular cognitive consort. My analysis of emotion is not meant to be taxonomically exhaustive. Only those features of emotion will be discussed which seem important in determining the relation of feeling to thought.

### I. OBJECTS OF EMOTION

For expository convenience I plan to stretch the label "emotion" to fit a heterogeneous assortment of reactions, moods, appetites, inclinations, aversions, desires, and attitudes, as well as emotions in the strict sense, like rage and disappointment. I mainly wish to exclude urges, impulses, decisions, intentions, resolutions, and similar dispositions to engage in some course of action. Consequently, to the extent that an emotion involves a proclivity to undertake overt action, as feeling vindictive includes a disposition to seek vengeance, I shall be silent about it.

In my discussion of thoughts, I shall concentrate upon opinions concerning matters of fact, as opposed to matters of value, which appear to be involved in emotions.<sup>1</sup> Thus, in examples (a) and (a'), I take it for granted that when John is embarrassed or worried that he is late for dinner, he thinks it is wrong, or at least embarrassing (worrisome) to arrive late for dinner. Surely John thinks one ought to be embarrassed (worried), or at least that it is normal to be ashamed (anxious) in the circumstances. I would find it odd for someone to be ashamed of a faultless performance, or to be worried about his good deeds. My reason for neglecting these normative beliefs is that their main function is to justify or excuse an emotion. By way of illustration, take a housewife who becomes infuriated when she is short-changed by the grocery clerk. She, or a bystander, might declare: "You have a right (it is perfectly natural) to get angry when you are short-changed." My analysis, in this case, would focus on the housewife's belief that the clerk actually cheated her; because this belief seems more closely related to her current emotion, her outrage with this clerk at this time, than her conviction that it is (generally) infuriating to be short-changed.

Now let me introduce some distinctions which will be crucial for an account of the relation between an emotion and the state of thought upon which the emotion is founded. All the emotions in my puzzling examples (a)–(c), as well as the straightforward (a')–(c'), share one feature which is not possessed by emotions like depression, euphoria, apathy, and the like: In each instance, John is worked up *about* something—his tardiness, the dessert wine, the hostess. As a contrast to his emotions, consider the following:

- (d) I am depressed this morning as a result of having consumed eight gin-fizzes last night; I neither believe nor doubt that I had gin-fizzes to drink. My memory of the previous night is a blank, because of alcohol or defense mechanisms. So I do not even wonder if I hit the bottle last night.
- (e) A young actress becomes euphoric when she learns that a maharajah wants to marry her. In fact, she becomes so ecstatic that she forgets all about the maharajah and his offer.
- (f) An art collector falls into a state of total apathy as a result of doubts concerning the authenticity of his favorite Pre-Columbian vases; ultimately he

loses his grip to the extent that he ceases to have any thoughts about his treasures and their origin.

In situation (d), I am not gloomy *about* anything; to say that I am depressed because I drank eight gin fizzes last night is only to specify the cause of my foul humor. How can I feel rotten over my excesses, since I have no idea that I committed them? With regard to situation (e), it would be incorrect—and linguistically odd—to say that the actress is euphoric about the maharajah's proposal, or euphoric that the maharajah wants her as a wife. How can she have any feelings about his offer, when she no longer believes or doubts that he has proposed to her? The art collector in (f) cannot be said to feel apathy toward his vases, because this phrase suggests, for one thing, that his mood is directed toward his pottery rather than his family, his business activities, politics, or what not; it also implies that he has some opinions regarding the vases. *Ex hypothesi*, neither of these implications hold: A man in a state of total apathy does not respond—even with indifference—to anything; furthermore, as I described the art collector, he has ceased to think about the vases.

The difference between (d)–(f) and each of the previous cases is that depression, euphoria, and total apathy have no objects. I shall explain this phrase in a moment; but first I want to say something about the connection between emotion and thought in (d)–(f). My gloom, in situation (d), would have been the effect of a belief, in case I had recalled my debauchery. Since I had no thoughts about my carousing, my gloom happens not to be a result of my cognitive state. The actress' euphoria, in story (e), happened to result from her antecedent belief that the maharajah wanted to marry her; still it is easy to imagine circumstances in which she becomes equally giddy without the concomitance of any beliefs; she might become euphoric as a result of severe fatigue before the movie-cameras. The same may be said of case (f): The collector's apathy might have resulted from a cerebral lesion, rather than doubts concerning the authenticity of his treasures. So (d)–(f) illustrate situations in which emotions and thoughts *are* related as effect to cause; they also prove that it is not self-contradictory to suppose that *some* emotions are unaccompanied by (not founded on) thoughts. I shall attempt to show, in Section IV, that this holds only for emotions without objects.

<sup>1</sup> In "Emotions," *Proceedings of the Aristotelian Society*, vol. 58 (1956–57), pp. 290–300, Mr. E. Bedford clearly demonstrates the connection between feelings and evaluations. For applications to moral philosophy, see G. Ryle, "Conscience and Moral Convictions," *Analysis*, vol. 7 (1940) and I. Thalberg, "Remorse" (forthcoming in *Mind*).

I said that depression, euphoria, and total apathy "have no objects." Grammar partly explains my meaning: We can say of someone that he is depressed, euphoric, or apathetic *simpliciter*; and there is no use for phrases like "So-and-so is depressed (euphoric, apathetic) that (about, with, at, over, on account of) . . ." Whereas, to start with (b'), (c), and (c'), we cannot say "John hopes," "John simply resents," or "John suspects." John must hope for something, or that something will or did happen; he must bear a grudge against, or feel misgivings toward, somebody. I need more than grammar to explain how embarrassment, worry, and anticipatory pleasure have objects; for it is quite correct to say, *tout court*, "John is embarrassed (worried, delighted)." However, it is a grammatical fact that there is a use for questions of the form, "What is he embarrassed (worried, delighted) about?" "Is he embarrassed (worried, delighted) that (with, etc.) . . .?" And this point leads to a more adequate explanation, namely that it is always germane to ask these questions if you are seeking an adequate characterization of a man's affective state.

We have, then, emotions which cannot take objects (depression, free-floating anxiety); emotions which must take objects (hope); and emotions which may be expected to have objects (embarrassment). Among the emotions which are always directed toward something, we may distinguish two sub-groups: Hate, aversion, love, admiration, appetite, and enjoyment must be focused on people, activities, events, things, or on groups of people, activities, events, or things. An infantryman cannot simply feel hate; he must, for instance, detest his sergeant or noncommissioned officers. A sybarite cannot simply feel aversion; he must dislike sports or physical exertion. I shall say that these emotions are directed toward *non-propositional* objects, to differentiate them from other "necessarily transitive" emotions whose objects may be characterized

by means of complete declarative sentences. Hope exemplifies the other subclass. I may hope (that) I have been appointed ambassador to Bolivia, and I may hope for the appointment or the ambassadorship. Of the emotions which *may* take objects, we seem to have a choice, in all cases, between propositional and non-propositional descriptions of their objects. Consider a unionist's feelings about the forthcoming strike. It is clearer to use the propositional idiom, and say, "He is glad (afraid, sorry, angry, dismayed) (that) there will be a strike"; but it is never incorrect to say, "He is pleased (worried, etc.) about the impending strike."<sup>2</sup>

Three points are worth noting in passing: (i) Object-expressions, propositional and non-propositional, do not always occupy the accusative position in a sentence which describes someone's emotion. A camper might say, "I shall come across a rattlesnake during my trip, which frightens me," rather than: "I am frightened that I shall come across a rattlesnake." Of course these sentences are not equivalent in meaning; the first one could be used to make a true statement only in case the speaker does meet a rattlesnake, whereas the camper may use the second sentence to give a correct description of his fear even if his trip does not include a confrontation with a rattler. (ii) Thus far, my characterization of the objects of emotion, and my distinction between propositional and non-propositional objects, are largely grammatical. Negatively, then, I have not attempted to say what sorts of things are designated by the expressions "his sergeant" and "there will be a strike," when these expressions appear in the sentences, "The infantryman hates his sergeant," and "The unionist is glad (that) there will be a strike." (iii) I have not discovered any philosophical significance in the grammatical fact, noted above, that emotions like hate and enjoyment are always given non-propositional objects. Nevertheless it is a fact that we

<sup>2</sup> Professor R. M. Chisholm makes use of the term "propositional object" in *Perceiving* (Ithaca, 1957), p. 142. It has been suggested to me (by Professor Mary Mothersill) that the objects of many emotions are best construed as possible or probable states of affairs. As I understand this view, it would have us say, of a prudent lathe-operator who buys accident-insurance, "He fears a possible (probable) injury," rather than "He fears an injury." In the *patois* of propositional objects, we should say, "He fears that he *might* be mutilated," or "He fears that he will probably be mutilated (that his chances of mutilation are such-and-such)," instead of "He fears that he will be mutilated." Is there any reason to adopt this analysis for fear and other anticipatory emotions? I think not. How can the object of the machinist's fear be a possible—or even a likely—injury? An *actual* mutilation will hurt; it will require hospital care; it will prevent the machinist from working; it will leave scars. If he suffers an actual mutilation, the additional fact that it was possible, or likely, will not make it any more painful, costly, time-consuming, or disfiguring. Further, imagine that the machinist retires without having suffered any actual injuries, although it was possible, or probable, that he would be injured. In other words, imagine that his injuries were merely possible or probable. Could he have feared these possible or probable, but nonactual, injuries? Did his possible or probable injuries hurt? Did they need medication? Did they keep him from working? Did they spoil his looks? If not, then probable or possible injuries were not what he feared when he bought insurance.

cannot specify a propositional object for the infantryman's loathing. Suppose, for example, that the sergeant has assigned our hero to Kitchen Police. We can say, "The soldier is furious (pleased, astonished) that the sergeant put him on K.P.," but not: "The soldier detests that the sergeant put him on K.P." And if we declare, "The soldier detests the sergeant because the sergeant put him on K.P.," we have not specified a propositional object of his loathing; rather, we have mentioned the fellow's grounds or reason for despising his sergeant.

## II. GROUNDS FOR EMOTION

The last distinction, between the objects of and the grounds for an emotion, demands some comment. Let me confess, to begin with, that two of my initial examples, (c) and (c'), are rather obviously mongrels: Does John feel resentment (misgivings) toward the hostess on the grounds that she gossiped about him? We can rid ourselves of an extra entity by saying: "He feels resentment about her having gossiped," and "He suspects that she has gossiped." In this sort of example, it is arbitrary to separate object and grounds. In other cases, the distinction is worse than arbitrary; it is logically untenable. Here is one illustration: The leader of a big-game expedition declares, "I'm frightened of cannibals because, after all, they do eat human flesh." There is an analytic connection between the phrases used to describe the objects of his fear (cannibals) and his reason for fearing them. Therefore no distinction is possible. Is the utterance of the safari leader too bizarre to prove that grounds are at times logically inseparable from objects of emotion? No, because his utterance does have a function which might not be served by the utterance, "I'm frightened of cannibals." The latter might leave us wondering: "What is it about cannibals that scares him? Do their war cries, their weapons, or their magical rites terrify him?" When he says he is frightened of them on account of their taste for human flesh, he has given us the information we need.

So much for the problematic cases. Let me illustrate how the distinction between *objects* of emotion and *grounds* for emotion may be free of the logical difficulties I noted above. For this purpose, consider a government security agent who learns that Brown, a candidate for the diplomatic service, receives a number of Marxist periodicals. The investigator begins to suspect that Brown has sub-

versive tendencies, because Brown subscribes to Marxist periodicals. However, the sleuth is not at all convinced that Brown is disloyal. Now plainly the security man does *not suspect* that Brown reads Marxist periodicals; his misgivings are directed toward Brown's loyalty, not Brown's reading habits; presumably the investigator is sure of Brown's reading habits, if that is why he suspects Brown of disloyalty. In this example, then, we must distinguish the object from the grounds. Perhaps I should append a *caveat*: To say, "The investigator has grounds (a reason) for suspecting Brown of subversive proclivities," is not to say that he has good grounds, epistemically or morally, to suspect Brown. The concept of "good grounds for emotion" will not be examined here.

We noted, in Section I, that some emotions, such as hate and enjoyment, must have objects. Do any require grounds as well? Surely I can like (or dislike) hillbilly songs "for no reason at all"; that is, I enjoy (abhor) them, but not because the tunes are catchy (simple-minded), or because the lyrics contain valuable information about American history (because the lyrics are stereotyped), and so forth. But can I desire revenge against my boss, or feel grateful toward him, and not desire revenge (feel gratitude) because of (for, on account of) anything he has done? Perhaps these are cases of "necessarily grounded" emotions; but the claim is very weak, since these examples do *not* involve a clear separation between the object of emotion and the grounds for emotion. The grounds disappear if, instead of saying, "I want to get revenge against my boss because he humiliated me," or "I'm grateful to him because he made me a junior partner of the firm," I say: "I want him to suffer the same humiliations he inflicted on me," or "I'm grateful that he made me a junior partner."

As long as it is clear that many emotions which must have objects need not have grounds, and that emotions which happen to have objects often do not have grounds, the distinction between object and grounds will advance our analysis of feeling and thought.

How are a man's thoughts connected with the object of, and the grounds for, his emotions? Let us start with a normal fellow, our cattle-breeder who believes that meat prices are falling, and who is indignant about it. From this description of him, it sounds as if the object of his indignation is his belief that meat prices are falling. But plainly this analysis of the object of his indignation will not work: For it is false to say, "He is indignant about his

belief that meat prices are falling," or "He is indignant that he believes that meat prices are falling." His outrage is not directed toward himself or his own belief! He is furious about prices.

What is the connection between his belief and the object of his emotion? His belief is directed toward the same object as his outrage; he believes that livestock prices are falling and he is indignant that they are.

This view, that objects of a man's emotion are objects of his thoughts, provides an answer to some questions the reader might have about my ontological commitments when I speak of emotions and their objects. Are emotions and their objects like automobiles and their engines, like teams and their coaches, like husbands and their wives, like citizens and their rights, like desks and their dimensions, like adjectives and the attributes they stand for? Any satisfactory reply<sup>8</sup> to the question, "What is the cattleman thinking about when he believes that meat prices are falling?" will be an answer to the metaphysical puzzle, "What kind of thing or event is the cattleman angry about when he is indignant that meat prices are falling?" If this phraseology sounds overly Platonic or Cartesian, then suppose that the cattleman is talking with his barber; he makes a bitter and abusive speech about the price situation; the object of his emotion turns out to be "what the cattleman was talking about (discussing, alluding, or referring to) in the course of his diatribe."

I shall not be able to dodge some other metaphysical questions, but let me postpone them in order to describe the relation between the grounds for an emotion and the thoughts (convictions, doubts, conjectures) upon which an emotion is founded. I shall use the cattleman again, and suppose that the *reason* he is indignant about falling beef-prices is that the Secretary of Agriculture had assured livestock owners that meat prices would be high this year. The cattleman believes that the Secretary made this claim. Under the circumstances it is correct to say, "The cattleman is outraged that meat prices are dropping,

because he believes that the Secretary of Agriculture assured. . . ." It would be pedantic to drive a logical wedge between the cattleman's belief and the grounds for his outrage. His belief is his reason for being angry. However, the belief upon which his emotion is *grounded* is not the belief upon which his emotion is *founded*: His indignation that prices are falling is *founded* on his belief that prices are falling; his indignation that prices are falling is *grounded* on his belief that the Secretary gave assurances to livestock owners. Is this distinction between "grounded" and "founded" a pedantic one? No! For it enables us to talk of groundless hopes, suspicions, and resentments, without implying that such emotions are unaccompanied by assumptions, surmises, or convictions. Suppose that our investigator, mentioned above, has no reason to suspect that Brown is unpatriotic. It would be true to say, "His suspicion is not grounded on any belief, conjecture, or assumption," but this does not mean: "The investigator neither believes, supposes, takes it for granted, conjectures, or doubts that Brown is unpatriotic!"

### III. ARE OBJECTS OF EMOTION CAUSES OF EMOTION?

My metaphysical worries would be over if I could identify objects of emotion with causes of emotion. In fact many of the things we like, fear, long for, and so on, are causes of our enjoyment, our fright, and our longing. If a rhinoceros is bearing down on me, and I fly in terror of it, the rhinoceros is both the object and the cause of my fear of it; you could use the same declarative sentence to describe the propositional object of my fear and the state of affairs which makes me frightened: "The rhinoceros is chasing him." (To say that the approaching rhinoceros causes my fear is not to deny that other circumstances—the condition of my nervous system, previous experiences with wild animals, etc.—contribute to my terror.)

Is it always possible to rank the object of emotion as a cause of emotion? Unfortunately for our meta-

<sup>8</sup> Dr. Richard M. Gale has an excellent, up-to-date review of the literature on "intentional objects," entitled "Propositions, Judgments, Sentences, and Statements," in the forthcoming *Encyclopedia of Philosophy* (Crowell-Collier, Free Press, and Macmillan Co.). His own doctrine is that we can believe or state what is not the case because verbs like "believe" and "state" "are used in such a way that there need not be anything in *rebus* answering to their grammatical accusatives"; he adds that "there is no need to introduce Platonic complexes, or anything else, to serve as the object" of thinking and asserting. I agree with this, but I wonder how we can—as we must—then provide for differences like the following: One girl is expecting a letter today, and another girl is expecting a telephone call; if neither of the girls receives what she expects, are we to say that they both have the same expectation, that they both expect nothing at all, or that they have no expectations? Shall we characterize the difference between their expectations in terms of the different descriptions they do or would give of what they expect ("a letter," "a telephone call")? This sounds like the most economical solution.

physics, no. For suppose that I am in a state of terror as I cross New York's Central Park, and when sympathetic passersby ask me what is bothering me, I declare: "I am terrified that I shall be chased by a rhinoceros during my stroll." Call my fear unmanly, ridiculous, or peculiar; the fact remains that the propositional object of my fear is correctly described by the sentence, "I shall be chased by a rhinoceros during my stroll." Does this sentence describe the cause of my fear? Well, perhaps a rhinoceros is in the offing—one escaped from the zoo and made its way to Central Park. In these circumstances it is true to say that I shall be chased by a rhinoceros. Even so, how could the future chase produce my current fear? In this case it is, to say the least, implausible to maintain that the object of my fear also causes my fear. And the assimilation of object to cause fails utterly when Central Park is free of rhinoceri. The sentence, "I shall be chased by a rhinoceros," still describes the propositional object of my fear, but does not, *ex hypothesi*, describe a cause of my fear. I assume that only actual events, states of affairs, and things can cause other events, states of affairs, and things. No rhinoceros is in the cards. Therefore it is false to say that my fear is caused by a rhinoceros attack. So the object of my emotion (a rhinoceros attack) is not a cause of my emotion.

Let me generalize from this odd example: Whenever an emotion is directed toward a future event or state of affairs, or something yet to be born, and whenever an emotion is concerned with something that did not, does not, or will not occur or exist, the object of emotion is not a cause of emotion. If very strict nominalists grumble—heroically—that a person *cannot really* fear what does not exist, my reply is this. Such a method of controlling the population of superfluous entities has the consequence that nobody has ever really feared ghosts, goblins, demons, or other imaginary powers. I doubt that superstitious fears will be destroyed by this sort of verbal magic.

#### IV. ARE THOUGHTS CAUSES OF EMOTION?

Our initial problem was to explain how emotions are based on thoughts (beliefs, assumptions, speculations). The main requirement we set was that the analysis of the relation should show whether it is possible, e.g., for a man to be embarrassed that he is late for dinner, but doubt that he is late for dinner. Enough distinctions have been established to consider one hypothesis: An emotion is founded

upon a thought if and only if the thought is a cause of the emotion. "Cause" must mean "empirically necessary condition" rather than "empirically sufficient condition." Even if my conviction that the milkman beats his horse sometimes makes me angry with the milkman, I may at times be too distracted with my own cares to work up any emotion about the milkman's cruelty, despite my belief that he is mean to his animal.

According to this hypothesis, the puzzling cases (a)–(c) (on the first page of this essay) would be at most causally impossible. I find some merit to the conclusion, but the analysis from which it follows is seriously defective.

If a man's thoughts are only causal conditions of his emotion, then it is conceivable that, for any emotion he has, he might have had precisely that emotion without the concurrence of any thoughts at all. On the causal analysis, people just happen to be fairly sure that some event will occur when they rejoice that it will occur. If their makeup or conditioning were different, they might be in the same affective state although they neither believe, doubt, nor imagine the event will occur.

Why do I resist this view? I admitted, in Section I, that moods like depression, euphoria, and total apathy may be caused by the subject's beliefs, doubts, or conjectures. Why not say the same of embarrassment and other emotions? Because the latter group of emotions may have objects, whereas depression, euphoria, total apathy, and free-floating anxiety—as its title indicates—are not about anything.

Let me explain why this fact invalidates the hypothesis that emotions with objects are effects of the convictions, doubts, or conjectures upon which these emotions are founded. Take the annoyance of a balletomane who thinks that no seats are left for this evening's dance recital. It is irrelevant, for the time being, to specify whether he simply takes it for granted that seats are unavailable, whether he is convinced of it, whether he merely conjectures that seats are unavailable, or whether he doubts it. Whatever the nature of his thought, could it be the cause of his annoyance that no seats are left? In so far as his annoyance is a form of agitation, including physiological disturbances and behavioral changes like fidgeting and inability to concentrate, his annoyance resembles objectless emotions. So his agitation might be a result of his thought that seats are unavailable. However, will any examination of his pulse rate, blood pressure, galvanic responses, salivation, or his bodily motions disclose

that he is annoyed over the lack of seats for tonight's performance? If not, then a study of his agitation will leave out an essential feature of his annoyance. It will not yield an answer to the question, always germane in such cases: "What is he annoyed about?"

Of course we shall get a quick answer if we interrogate him. His reply, "I'm annoyed that no tickets are left," indicates the propositional object of his chagrin. But his words also reveal that he has thoughts regarding the tickets. How does this prove that his thought cannot be a cause of his annoyance that tickets are gone? It seems to me that any time you claim one event or condition is a cause of another event or condition, you must be able to gather evidence of the effect which is logically independent of your evidence of its putative cause. According to this principle, you may claim, "There was a short circuit in the warehouse, which caused a fire in the warehouse," but not: "There were flames in the warehouse, which caused a fire in the warehouse." Evidence of a conflagration is not always evidence of a short circuit, but it is always evidence of flames. How does this principle apply to the particular hypothesis concerning the ballet enthusiast, "His thought that tickets are gone caused his annoyance that tickets are gone"? Well, it appears that if we prove he is vexed that tickets are gone, we also prove that he thinks (believes, conjectures, doubts) that tickets are gone; therefore we cannot claim that his emotion is the effect of his thought.

In the foregoing argument I took it for granted that only a man's verbal behavior will give us evidence regarding the object of his emotion. Will my conclusion, that emotions are not the effects of the thoughts upon which they are founded, hold without this methodological assumption? Since the behavior of chimpanzees and other speechless

animals permits us to ascribe rage, fear, and dejection to them, perhaps verbal behavior is dispensable in the case of human beings. So consider, for purposes of the argument, some nonverbal responses and activities which might be evidence for saying that our ballet enthusiast is annoyed about the lack of tickets: An hour before curtain time, he approaches the box office, which is closed; his blood pressure rises, his face reddens, he bites his lip and stomps off toward the nearest cowboy film. Now we cannot grant that he *read* the sign on the box office, "Tonight's performance sold out," because this would automatically show he had thoughts regarding the lack of tickets. In the circumstances, what would entitle us to assert, "He is vexed that tickets are gone," rather than any or all of the following: "He is vexed that the box office is shut," "He is vexed that he made a fruitless visit to the box office," or "He is vexed"? It seems to me that any other clues about the object of his annoyance would also indicate that he had thoughts regarding the same object.

I would give a similar interpretation of the non-human cases on which this objection depends. Take a psychologist's rats: They are conditioned to display fear-reactions whenever they are placed in a compartment where they had received electric shocks. Now we put them in their chamber of horrors. They dash about, squealing pathetically, even though they are not being shocked this time. If their agitation entitles us to say, "Now they fear they will be shocked," then do we have to perform additional experiments before we can claim, "They think they will be shocked"? I do not claim that any psychologist would say, "The rats fear they will be shocked"; but if he did demonstrate this claim, it seems to me he would thereby demonstrate that his rats think they will be shocked. I conclude that the relation between thought and emotions which have objects is not a causal relation.<sup>4</sup>

<sup>4</sup> The causal theory is not entirely made of straw; it is Hume's doctrine in the *Treatise*. See pages 368, 386, 415-426, 439-446, Selby-Bigge edition (Oxford, 1955). His kind of analysis of the relation between feeling and thought has been under intermittent attack, but I find the most current objections fairly weak. It is said that my belief that I inherited somebody's fortune cannot be a cause of my pleasure that I did, because: (i) I cannot be mistaken when I relate my pleasure to my belief; however, one can, in principle, be mistaken regarding the causes of any phenomenon; (ii) I never need to observe the concomitance of my belief and my pleasure, nor do I make use of inductive evidence of similar affective-cognitive couples in my past experience; but one always needs this sort of evidence when claiming that something is the cause of another thing; (iii) thoughts and many emotions are not processes, but causes and their effects are events having earlier and later spatio-temporal phases. Of course I accept the conclusion of these arguments, but I think that (i) confuses two distinct notions: a person's *de facto* or (perhaps) *de jure* incorrigibility when he makes some first-person psychological statements, and the logical status of assertions (by anyone) regarding someone's psychological condition; (ii) seems to me to embody a similar confusion, between the plausible claim that one does not utilize observational procedures in making some first-person psychological statements, and the dubious claim that inductive evidence is irrelevant to other people's assessment of the truth or falsity of one's first-person statements; (iii) seems to me to depend on a very restrictive notion of "cause and effect." For extremely persuasive developments of (i), (ii), (iii), and other arguments against the casual view, consult: B. A. O. Williams, "Pleasure and Belief," *Proceedings of the Aristotelian Society Supplementary Volume* 33 (1959), pp. 57-59; and J. Teichmann, "Mental Cause and Effect," *Mind*, vol. 70 (1961), January number.

## V. TWO ALTERNATIVES

What have we proved thus far? Only that emotions which have objects are logically tied to some form of thought about the same object. So there are two methods of resolving the odd cases I listed at the beginning of this paper:

(1) We can declare that each sort of emotion which has an object is based (founded) upon a particular kind of thought, in the sense that the latter is a logically necessary condition of the former;<sup>5</sup>

(2) We can hold that each sort of emotion is based upon some form of thought or other, i.e., thought is a logically necessary condition of every emotion which has an object, but it happens that people have some kinds of emotion when they doubt, and other kinds when they conjecture.

With regard to example

(a) John is embarrassed that he is late for dinner, but he doubts that he is (late for dinner),

a champion of the first analysis would assert that my imaginary situation is logically impossible: My description of John's mental state is self-contradictory. According to the second view, the situation is unusual though possible; my description of John is unlikely to be true, but it might be. Analysis (1) commits us to the view that the sentential function, "— is embarrassed that . . ." means, *inter alia*, "— believes that . . ."; "— is worried that . . ." means, *inter alia*, "— believes or conjectures or doubts that . . ."; and so on for each item in our emotion-vocabulary. Analysis (2) requires, for its intelligibility, some hint of an explanation for the observed concomitance of types of emotion and types of thought, if this correlation is not due to the meaning of emotion-words.

A number of considerations favor analysis (1). The argument I used against the hypotheses that there is only a causal relation between emotion and thought applies here with equal force. Any proof that John is embarrassed that he is late for dinner seems to prove that John is convinced he is late, and thus to falsify the conjunction, "John is embarrassed that he is late for dinner, but doubts that he is late for dinner." If John makes a sincere avowal, "I'm embarrassed that I'm late for dinner," has he not expressed his conviction that he is late? Plainly he has not manifested disbelief! Add to this evidence

various forms of non-linguistic behavior: his manner, his gestures and carriage, his sheepish grin, his blushes. Isn't this the picture of a man who believes he is late?

Now what if all this confirms the seemingly redundant hypothesis, "John is embarrassed that he is late, and he believes he is late"? Have we thereby refuted the claim, "John is embarrassed that he is late, but he doubts he is late"? To save space, I wish to use shorthand names for the statements under examination:

*E* for the statement "John is embarrassed that he is late"

*D* for "John doubts he is late"

*B* for "John believes he is late"

*B'* for "John believes he is not late"

*not-B* for "It is not the case that John believes he is late"

If we use "&" as a name of the conjunction operator, then example (a) will read: *E* & *D*. My problem is this: I admit that any proof of *E* will count in favor of *B*. Therefore I must admit that nothing would ever prove that *E* & *D* is true.

I think the argument is faulty. It does not matter whether *E* entails *B*, or whether the truth of *E* makes *B* very probable. It is not obvious to me that the truth of *B* entails the falsity of *D*. I assume that *D* is equivalent to *B'* rather than *not-B*. If that is so, then the truth of *E* & *B* does not entail the falsity of *E* & *B'* (*E* & *D*), although *E* & *B* does entail the falsity of *E* & *not-B*. In other words, *E*, *B* and *B'* might all be true. That is, John might believe he is late and believe he is not late.

Is such a thing logically possible? I admit the following points: (i) John's belief that he is late is inconsistent with his belief that he is not late; (ii) consequently, to assert "John believes he is late and John believes he is not late" (*B* & *B'*) is to assert that John has inconsistent beliefs. Notice, however, that (i) and (ii) do not show that the statement *B* & *B'* is an inconsistent statement.

At this juncture, a partisan of analysis (1) will reformulate his argument along the following lines: "Of course it would beg the question to say that nobody can (logically) hold inconsistent beliefs; so *B* & *B'* is not self-contradictory; however, what evidence could possibly show that *B* and *B'* are conjointly true? Whatever counts in favor of *B* automatically discredits *B'* to the same degree that

<sup>5</sup> This is Bedford's view in "Emotions," cited above. Discussing the statement, "A resents what B did," Bedford writes: "A's belief that B has done something that affects him adversely is . . . a necessary condition if the word 'resentment' is to be used at all" (p. 295); elsewhere he says, "The expression 'I hope that . . .' implies . . . a very vague estimate of probability . . ." (p. 293).

it favors *B*, and *vice versa*." If I accept this last methodological premise, I would be guilty of unverifiable speculation if I said, "*E* is true, so *B* is (probably) true; but perhaps *B'* (i.e., *D*) is true as well."

It seems to me that this argument begs the same question in methodological disguise: Whatever entitles us to say that evidence for *B* counts against *B'*? Take some evidence for *B*: Another guest at the dinner-party asks John, "Do you believe you are late?" and John replies affirmatively. If this disconfirms *B'* to the same degree that it confirms *B*, then we must have assumed that a person cannot (logically) believe and disbelieve the same thing. And consider some evidence for *B'*: John's wife telephones him to ask whether he arrived on time, and he tells her: "It was a wild trip, but I believe that I'm not late." If we take this as a sincere avowal of his belief, his declaration will have equal weight against statement *B* only on the assumption that *B'* and *B* are logically incompatible.<sup>6</sup>

The upshot is that analysis (1) will exclude the imaginary situations (a)–(c) only if we take it for granted that nobody can (logically) hold inconsistent beliefs. It is not my purpose here either to vindicate or to give a counter-example to that difficult assumption. However, I should like to show that analysis (2) will take care of the odd cases without the disputed assumption. Let me explain this alternative analysis by reference to,

(b) John is delighted that there will be champagne with dessert, but he merely conjectures that there will be.

Why does this statement bother us, although its counterpart,

(b') John hopes that there will be champagne with dessert, but he merely conjectures that there will be,

sounds quite plausible?

A follower of the second view would explain that in (b), John's emotion is somehow inappropriate to his state of thought; there is an incongruity between his conjecture and his delight. (b'), on the other hand, ascribes an emotion to John which somehow fits his cognitive outlook. I suggest two interpretations of the claim that a conjecture is an inappropriate foundation for anticipatory pleasure, but an

appropriate basis for hope. The verdict, "His emotion is inappropriate," may mean:

(I) It is not normal for people to feel pleasure about a future event when they merely conjecture that it will occur. People do not, as a rule, feel that sort of emotion, or that degree of it, when they are only guessing; so on statistical grounds alone we wonder: "Perhaps John is secretly convinced that champagne will be served? Or if he is just guessing, then he is probably in a state of hope rather than anticipatory delight." Statistics will entitle us to assume that John's thoughts are consistent, i.e., that he does not believe or doubt that champagne is forthcoming at the same time that he conjectures that champagne is forthcoming. We can go on to account for the concomitant variation we observe in John's community between people's emotions and their states of thought, as well as their tendency to alter their beliefs when they notice inconsistencies among their opinions. I assume that the account would show the observed regularities to be the effect of brain processes, social conditioning, and so forth. If we have an adequate explanation of the regularities, it would also indicate how odd cases like (b) might result from cerebral lesions, childhood traumas, or insufficient training. There is no need to add the following: We might find quite different correlations between types of emotion and types of thought, when we compare groups which have diverse hereditary proclivities or dissimilar social institutions.

(II) A man's emotions may be inappropriate—in kind or degree—to his state of thought in another respect. We might judge that John is unreasonable to feel anticipatory pleasure when he merely conjectures that there will be champagne. Is it reasonable for him to hope? At least it is not unreasonable. The notion that our feelings can be reasonable or unreasonable, *vis-à-vis* our cognitive state, depends upon their similarity to many of our habitual and instinctive actions. It is reasonable for a pugilist to duck or to raise his guard when he believes his opponent is throwing an uppercut, and it is unwise of him to stiffen before the other man's punch. To be sure, the boxer does not decide to move out of range or to hold his position; but his movements are reasonable or unreasonable in the sense that they are effective means for him to achieve victory and to avoid a

<sup>6</sup> The argument in this and the four preceding paragraphs owes whatever merit it possesses to comments I received from Professors Sue Larsen (Stanford University) and Roderick M. Chisholm (Brown University). My main reason for supposing it is logically possible for people to have inconsistent beliefs is that people are urged and trained to be consistent and blamed when they appear to hold incompatible opinions. The fact that a man with inconsistent beliefs has, in most societies, very little chance of persuading others to adopt his opinions is a strong inducement for him to avoid self-contradiction.

pummeling. Now compare emotions like anticipatory pleasure to the boxer's defensive activity. Emotions of this sort are also forms of readiness. In example (b), John is prepared for champagne. Perhaps he is *set to drink* champagne as well, but we are not considering the conative aspects of emotion. How then may emotions, as states of readiness, acquire the labels "reasonable" and "unreasonable"? A man's emotions are reasonable when, in view of the man's beliefs, doubts, or conjectures, the form of readiness they involve is likely to be effective and necessary. When they involve inadequate or superfluous preparations, his feelings are unreasonable. In example (b), John only guesses there will be champagne; consequently he must think that his delight is probably an otiose form of preparation. He is like a boxer who engages in elaborate dodging and protective maneuvers when he believes his challenger is staggering on the ropes.

Fear, as usual, is a perspicacious model for reasonable and unreasonable emotion. If your friend's train is several hours overdue, which leads you to believe it was derailed, it is reasonable for you to worry that his train was in a mishap. If the wreck is suddenly announced, you will be less shaken than if you had not prepared yourself. What if his train is not behind schedule at all, but you are struck by the thought that it crashed? Then you stir yourself up needlessly if you worry. Your anguish is unreasonable—not because it is unlikely to be "fulfilled," not because you will be "disappointed," but because your agonizing preparations are unlikely to serve their function in this case. Perhaps you cannot help worrying anyway, but that only proves you are not to blame for your unreasonable fears.

This comparison between emotions and preparatory activity only makes sense if we assume that both have a function in the life of a reasonable person. In saying that a man's actions or feelings are reasonable, we presuppose that he thinks rationally; when he is certain of something, it is more likely to be true than when he has a hunch about it, or doubts it; and we presuppose that he sets some value upon the ends that are to be achieved by preparatory activities and emotions. We assume, for

instance, when the boxer *believes* his opponent is aiming an uppercut at him, an uppercut is more likely than when the boxer doubts or conjectures that his opponent is loosing an uppercut. We also take it for granted that our pugilist wants to defeat his opponent and to avoid his blows.

A satisfactory analysis of these assumptions, which are packed into the concept of reasonable and unreasonable emotion, would take us far beyond the topic of this paper, "How are a man's worries, hopes, resentments, and so on related to his states of thought?" In particular, the claim that a person's emotion is not just statistically odd, but unreasonable, raises problems about the justification of emotions. Evidently, to say of John in example (b), "He is unreasonable to be delighted that there will be champagne, when he only has a hunch that champagne is on the way," is to suggest that his state of thought does not justify his emotion; he should not feel that way on the basis of a guess. I cannot undertake an account of these normative concepts here.<sup>7</sup>

My thesis in this final section was that two analyses of the rapport between feelings and thought will do justice to the important distinctions made in the preliminary discussion of objects, grounds, and causes of emotion. In section IV we had to discard the view that a person's beliefs, doubts, and speculations are nothing but causes of his emotions; we rejected it, ultimately, because it was epistemologically vacuous, for we could discover no way to specify the object of someone's emotion without assuming that he had thoughts about the object. The remaining views may be reformulated quite simply:

(1) A particular type of thought is a logically necessary condition of each type of emotion which has an object;

(2) Some type of thought is a logically necessary condition of each type of emotion which has an object, and each type of emotion is appropriate to particular types of thought, inappropriate to others.

If we take (1) as an explication of phrases like "resignation is founded on certainty, pleasure is founded on belief," the relation between emotions

<sup>7</sup> Aristotle seems to take the view that emotions can be appropriate, in the sense that they may serve important functions in the life of a reasonable man. See *Nicomachean Ethics* ii 1106b 15-25; also Mrs. Warnock and Professor Ewing, "The Justification of Emotions," *Proc. Arist. Soc. Suppl. Vol.* 31 (1957). The appropriateness relation is more evident when we consider conative states. Aristotle declares that intention or choice "cannot relate to impossibles, and if anyone said he chose them he would be thought silly; but there may be a wish even for impossibles" (op. cit., iii 1111b 20-25). Is it logically possible for a man to be "silly" in this way—to decide or intend to do something he believes he cannot do? Such a man would be unreasonable, of course, and statistically rare to boot. See my discussion of "Intending the Impossible," *Australasian Journal of Philosophy*, vol. 40 (1962), May number.

and their cognitive cohorts is entirely logical. However, we can exclude the difficult cases (a)–(c) only by appealing to a disputable maxim, “A person cannot (logically) hold inconsistent beliefs.” (2) makes the relation between emotion and thought partly logical, partly causal, and—if we say that “appropriate” means “reasonable” as well as “statistically normal”—partly normative. Analysis (2) seems more in keeping with a flexible, empirically oriented psychological theory of emotion. (1) would present us with fixed *a priori* relations between various forms of emotion and thought.

Explication (2) of what it is for an emotion to be founded upon an appropriate type of thought does appeal to some obscure normative considerations. But this is not a compelling reason to discard (2). In fact, it may be illuminating to investigate further the similarities, (i) between admiring, disliking, resenting, or fearing something, and appraising or evaluating it; (ii) between moral justification or censure of someone's emotion, and the kind of assessment that is involved in judging someone's emotion appropriate or inappropriate to his beliefs, doubts, or conjectures.<sup>8</sup>

<sup>8</sup> I am indebted to a number of people, either for their views on the problems I have discussed here, or for their comments on earlier drafts of this paper. In particular I wish to thank Professors Mary Mothersill, Sue Larsen, Terence Penelhum, Roderick M. Chisholm, and an anonymous referee for the *American Philosophical Quarterly*.

*University of Washington*

## V. THE MYTH OF THE AESTHETIC ATTITUDE

GEORGE DICKIE\*

SOME recent articles<sup>1</sup> have suggested the unsatisfactoriness of the notion of the aesthetic attitude and it is now time for a fresh look at that encrusted article of faith. This conception has been valuable to aesthetics and criticism in helping wean them from a sole concern with beauty and related notions.<sup>2</sup> However, I shall argue that the aesthetic attitude is a myth and while, as G. Ryle has said, "Myths often do a lot of theoretical good while they are still new,"<sup>3</sup> this particular one is no longer useful and in fact misleads aesthetic theory.

There is a range of theories which differ according to how strongly the aesthetic attitude is characterized. This variation is reflected in the language the theories employ. The strongest variety is Edward Bullough's theory of psychical distance, recently defended by Sheila Dawson.<sup>4</sup> The central technical term of this theory is "distance" used as a verb to denote an action which either constitutes or is necessary for the aesthetic attitude. These theorists use such sentences as "He distanced (or failed to distance) the play." The second variety is widely held but has been defended most vigorously in recent years by Jerome Stolnitz and Eliseo Vivas. The central technical term of this variety is "disinterested"<sup>5</sup> used either as an adverb or as an adjective. This weaker theory speaks not of a special kind of action (distancing) but of an ordinary kind of action (attending) done in a certain way (disinterestedly). These first two versions are perhaps not as different as my classification suggests. How-

ever, the language of the two is different enough to justify separate discussions. My discussion of this second variety will for the most part make use of Jerome Stolnitz' book<sup>6</sup> which is a thorough, consistent, and large-scale version of the attitude theory. The weakest version of the attitude theory can be found in Vincent Tomas' statement "If looking at a picture and attending closely to how it looks is not really to be in the aesthetic attitude, then what on earth is?"<sup>7</sup> In the following I shall be concerned with the notion of *aesthetic* attitude and this notion may have little or no connection with the ordinary notion of an *attitude*.

### I

Psychical distance, according to Bullough, is a psychological process by virtue of which a person *puts* some object (be it a painting, a play, or a dangerous fog at sea) "out of gear" with the practical interests of the self. Miss Dawson maintains that it is "the beauty of the phenomenon, which captures our attention, puts us out of gear with practical life, and forces us, if we are receptive, to view it on the level of aesthetic consciousness."<sup>8</sup>

Later she maintains that some persons (critics, actors, members of an orchestra, and the like) "distance deliberately."<sup>9</sup> Miss Dawson, following Bullough, discusses cases in which people are unable to bring off an act of distancing or are incapable of being induced into a state of being

\*I wish to thank both Monroe C. Beardsley and Jerome Stolnitz who read earlier drafts of this paper and made many helpful comments.

<sup>1</sup> See Marshall Cohen, "Appearance and the Aesthetic Attitude," *Journal of Philosophy*, vol. 56 (1959), p. 926; and Joseph Margolis, "Aesthetic Perception," *Journal of Aesthetics and Art Criticism*, vol. 19 (1960), p. 211. Margolis gives an argument, but it is so compact as to be at best only suggestive.

<sup>2</sup> Jerome Stolnitz, "Some Questions Concerning Aesthetic Perception," *Philosophy and Phenomenological Research*, vol. 22 (1961), p. 69.

<sup>3</sup> *The Concept of Mind* (London, 1949), p. 23.

<sup>4</sup> "Distancing" as an Aesthetic Principle," *Australasian Journal of Philosophy*, vol. 39 (1961), pp. 155-174.

<sup>5</sup> "Disinterested" is Stolnitz' term. Vivas uses "intransitive."

<sup>6</sup> *Aesthetics and Philosophy of Art Criticism* (Boston, 1960), p. 510.

<sup>7</sup> "Aesthetic Vision," *The Philosophical Review*, vol. 68 (1959), p. 63. I shall ignore Tomas' attempt to distinguish between appearance and reality since it seems to confuse rather than clarify aesthetic theory. See F. Sibley, "Aesthetics and the Looks of Things," *Journal of Philosophy*, vol. 56 (1959), pp. 905-915; M. Cohen, op. cit., pp. 915-926; and J. Stolnitz, "Some Questions Concerning Aesthetic Perception," op. cit., pp. 69-87. Tomas discusses only visual art and the aesthetic attitude, but his remarks could be generalized into a comprehensive theory.

<sup>8</sup> Dawson, op. cit., p. 158.

<sup>9</sup> Ibid., pp. 159-160.

distanced. She uses Bullough's example of the jealous ("under-distanced") husband at a performance of *Othello* who is unable to keep his attention on the play because he keeps thinking of his own wife's suspicious behavior. On the other hand, if "we are mainly concerned with the technical details of its [the play's] presentation, then we are said to be over-distanced."<sup>10</sup> There is, then, a species of action—distancing—which may be deliberately done and which initiates a state of consciousness—being distanced.

The question is: Are there actions denoted by "to distance" or states of consciousness denoted by "being distanced"? When the curtain goes up, when we walk up to a painting, or when we look at a sunset are we ever induced into a state of being distanced either by being struck by the beauty of the object or by pulling off an act of distancing? I do not recall committing any such special actions or of being induced into any special state, and I have no reason to suspect that I am atypical in this respect. The distance-theorist may perhaps ask, "But are you not usually oblivious to noises and sights other than those of the play or to the marks on the wall around the painting?" The answer is of course—"Yes." But if "to distance" and "being distanced" simply mean that one's attention is focused, what is the point of introducing new technical terms and speaking as if these terms refer to special kinds of acts and states of consciousness? The distance-theorist might argue further, "But surely you put the play (painting, sunset) 'out of gear' with your practical interests?" This question seems to me to be a very odd way of asking (by employing the technical metaphor "out of gear") if I attended to the play rather than thought about my wife or wondered how they managed to move the scenery about. Why not ask me straight out if I paid attention? Thus, when Miss Dawson says that the jealous husband under-distanced *Othello* and that the person with a consuming interest in techniques of stagecraft over-distanced the play, these are just technical and misleading ways of describing two different cases of inattention. In both cases something is being attended to, but in neither case is it the action of the play. To introduce the technical terms—"distance," "under-distance," and "over-distance" does nothing but send us chasing after phantom acts and states of consciousness.

Miss Dawson's commitment to the theory of distance (as a kind of mental insulation material

necessary for a work of art if it is to be enjoyed aesthetically) leads her to draw a conclusion so curious as to throw suspicion on the theory.

One remembers the horrible loss of distance in *Peter Pan*—the moment when Peter says "Do you believe in fairies? . . . If you believe, clap your hands!" the moment when most children would like to slink out of the theatre and not a few cry—not because Tinkerbell may die, but because the magic is gone. What, after all, should we feel like if Lear were to leave Cordelia; come to the front of the stage and say, "All the grown-ups who think that she loves me, shout 'Yes'!"<sup>11</sup>

It is hard to believe that the responses of any children could be as theory-bound as those Miss Dawson describes. In fact, *Peter Pan*'s request for applause is a dramatic high point to which children respond enthusiastically. The playwright gives the children a momentary chance to become actors in the play. The children do not at that moment lose or snap out of a state of being distanced because they never had or were in any such thing to begin with. The comparison of *Peter Pan*'s appeal to the hypothetical one by Lear is pointless. *Peter Pan* is a magical play in which almost anything can happen, but *King Lear* is a play of a different kind. There are, by the way, many plays in which an actor directly addresses the audience (*Our Town*, *The Marriage Broker*, *A Taste of Honey*, for example) without causing the play to be less valuable. Such plays are unusual, but what is unusual is not necessarily bad; there is no point in trying to lay down rules to which every play must conform independently of the kind of play it is.

It is perhaps worth noting that Susanne Langer reports the reaction she had as a child to this scene in *Peter Pan*.<sup>12</sup> As she remembers it, *Peter Pan*'s appeal shattered the illusion and caused her acute misery. However, she reports that all the other children clapped and laughed and enjoyed themselves.

## II

The second way of conceiving of the aesthetic attitude—as the ordinary action of attending done in a certain way (disinterestedly)—is illustrated by the work of Jerome Stolnitz and Eliseo Vivas. Stolnitz defines "aesthetic attitude" as "disinterested and sympathetic attention to and contemplation of any object of awareness whatever, for its own sake alone."<sup>13</sup> Stolnitz defines the main

<sup>10</sup> Ibid., p. 159.

<sup>11</sup> Ibid., p. 168.

<sup>12</sup> *Aesthetics and Philosophy of Art Criticism*, pp. 34–35.

<sup>13</sup> *Feeling and Form* (New York, 1953), p. 318.

terms of his definition: "disinterested" means "no concern for any ulterior purpose";<sup>14</sup> "sympathetic" means "accept the object on its own terms to appreciate it";<sup>15</sup> and "contemplation" means "perception directed toward the object in its own right and the spectator is not concerned to analyze it or ask questions about it."<sup>16</sup>

The notion of disinterestedness, which Stolnitz has elsewhere shown<sup>17</sup> to be seminal for modern aesthetic theory, is the key term here. Thus, it is necessary to be clear about the nature of disinterested attention to the various arts. It can make sense to speak, for example, of listening disinterestedly to music only if it makes sense to speak of listening interestedly to music. It would make no sense to speak of walking *fast* unless walking could be done *slowly*. Using Stolnitz' definition of "disinterestedness," the two situations would have to be described as "listening with no ulterior purpose" (disinterestedly) and "listening with an ulterior purpose" (interestedly). Note that what initially appears to be a perceptual distinction—listening in a certain way (interestedly or disinterestedly)—turns out to be a motivational or an intentional distinction—listening for or with a certain purpose. Suppose Jones listens to a piece of music for the purpose of being able to analyze and describe it on an examination the next day and Smith listens to the same music with no such ulterior purpose. There is certainly a difference between the motives and intentions of the two men: Jones has an ulterior purpose and Smith does not, but this does not mean Jones's *listening* differs from Smith's. It is possible that both men enjoy the music or that both be bored. The attention of either or both may flag and so on. It is important to note that a person's motive or intention is different from his action (Jones's listening to the music, for example). There is only one way to *listen* to (to attend to) music, although the listening may be more or less attentive and there may be a variety of motives, intentions, and reasons for doing so and a variety of ways of being distracted from the music.

In order to avoid a common mistake of aestheticians—drawing a conclusion about one kind of art and assuming it holds for all the arts—the question of disinterested attention must be considered for arts other than music. How would one look at a

painting disinterestedly or interestedly? An example of alleged interested viewing might be the case in which a painting reminds Jones of his grandfather and Jones proceeds to muse about or to regale a companion with tales of his grandfather's pioneer exploits. Such incidents would be characterized by attitude-theorists as examples of using a work of art as a vehicle for associations and so on, i.e., cases of interested attention. But Jones is not looking at (attending to) the painting at all, although he may be facing it with his eyes open. Jones is now musing or attending to the story he is telling, although he had to look at the painting at first to notice that it resembled his grandfather. Jones is not now looking at the painting interestedly, since he is not now looking at (attending to) the painting. Jones's thinking or telling a story about his grandfather is no more a part of the painting than his speculating about the artist's intentions is and, hence, his musing, telling, speculating, and so on cannot properly be described as attending to the painting interestedly. What attitude-aestheticians are calling attention to is the occurrence of irrelevant associations which distract the viewer from the painting or whatever. But distraction is not a special kind of attention, it is a kind of inattention.

Consider now disinterestedness and plays. I shall make use of some interesting examples offered by J. O. Urmson,<sup>18</sup> but I am not claiming that Urmson is an attitude-theorist. Urmson never speaks in his article of aesthetic attitude but rather of aesthetic satisfaction. In addition to aesthetic satisfaction, Urmson mentions economic, moral, personal, and intellectual satisfactions. I think the attitude-theorist would consider these last four kinds of satisfaction as "ulterior purposes" and, hence, cases of interested attention. Urmson considers the case of a man in the audience of a play who is delighted.<sup>19</sup> It is discovered that his delight is *solely* the result of the fact that there is a full house—the man is the impresario of the production. Urmson is right in calling *this* impresario's satisfaction economic rather than aesthetic, although there is a certain oddness about the example as it finds the impresario sitting *in the audience*. However, my concern is not with Urmson's examples as such but with the attitude theory. This impresario is certainly an interested party in the fullest sense of the word, but is his

<sup>14</sup> *Ibid.*, p. 35.

<sup>15</sup> *Ibid.*, p. 36.

<sup>16</sup> *Ibid.*, p. 38.

<sup>17</sup> "On the Origins of 'Aesthetic Disinterestedness,'" *The Journal of Aesthetics and Art Criticism*, vol. 20 (1961), pp. 131-143.

<sup>18</sup> "What Makes a Situation Aesthetic?" in *Philosophy Looks at the Arts*, Joseph Margolis (ed.), (New York, 1962). Reprinted from *Proceedings of the Aristotelian Society, Supplementary Volume 31* (1957), pp. 75-92.

<sup>19</sup> *Ibid.*, p. 15.

behavior an instance of interested attention as distinct from the supposed disinterested attention of the average citizen who sits beside him? In the situation as described by Urmson it would not make any sense to say that the impresario is attending to the play at all, since his *sole* concern at the moment is the till. If he can be said to be attending to anything (rather than just thinking about it) it is the size of the house. I do not mean to suggest that an impresario could not attend to his play if he found himself taking up a seat in a full house; I am challenging the sense of disinterested attention. As an example of personal satisfaction Urmson mentions the spectator whose daughter is in the play. Intellectual satisfaction involves the solution of technical problems of plays and moral satisfaction the consideration of the effects of the play on the viewer's conduct. All three of these candidates which the attitude-theorist would propose as cases of interested attention turn out to be just different ways of being distracted from the play and, hence, not cases of interested attention to the play. Of course, there is no reason to think that in any of these cases the distraction or inattention must be total, although it could be. In fact, such inattentions often occur but are so fleeting that nothing of the play, music, or whatever is missed or lost.

The example of a playwright watching a rehearsal or an out-of-town performance with a view to rewriting the script has been suggested to me as a case in which a spectator is certainly attending to the play (unlike our impresario) and attending in an interested manner. This case is unlike those just discussed but is similar to the earlier case of Jones (not Smith) listening to a particular piece of music. Our playwright—like Jones, who was to be examined on the music—has ulterior motives. Furthermore, the playwright, unlike an ordinary spectator, can change the script after the performance or during a rehearsal. But how is our playwright's *attention* (as distinguished from his motives and intentions) different from that of an ordinary viewer? The playwright might enjoy or be bored by the performance as any spectator might be. The playwright's attention might even flag. In short, the kinds of things which may happen to the playwright's attention are no different from those that may happen to an ordinary spectator,

although the two may have quite different motives and intentions.

For the discussion of disinterested-interested reading of literature it is appropriate to turn to the arguments of Eliseo Vivas whose work is largely concerned with literature. Vivas remarks that "By approaching a poem in a nonaesthetic mode it may function as history, as social criticism, as diagnostic evidence of the author's neuroses, and in an indefinite number of other ways."<sup>20</sup> Vivas further notes that according to Plato "the Greeks used Homer as an authority on war and almost anything under the sun," and that a certain poem "can be read as erotic poetry or as an account of a mystical experience."<sup>21</sup> The difference between reading a poem *as* history or whatever (reading it nonaesthetically) and reading it aesthetically depends on how *we* approach or read it. A poem "does not come self-labelled,"<sup>22</sup> but presumably is a poem only when it is read in a certain way—when it is an object of aesthetic experience. For Vivas, being an aesthetic object means being the object of the aesthetic attitude. He defines the aesthetic experience as "an experience of rapt attention which involves the intransitive apprehension of an object's immanent meanings and values in their full presentational immediacy."<sup>23</sup> Vivas maintains that his definition "helps me understand better what I can and what I cannot do when I read *The Brothers [Karamazov]*" and his definition "forces us to acknowledge that *The Brothers Karamazov* can hardly be read as art. . . ."<sup>24</sup> This acknowledgment means that we probably cannot intransitively apprehend *The Brothers* because of its size and complexity.

"Intransitive" is the key term here and Vivas' meaning must be made clear. A number of passages reveal his meaning but perhaps the following is the best. "Having once seen a hockey game in slow motion, I am prepared to testify that it was an object of pure intransitive experience [attention]—for I was not *interested* in which team won the game and no external factors mingled with my interest in the beautiful rhythmic flow of the slow-moving men."<sup>25</sup> It appears that Vivas' "intransitive attention" has the same meaning as Stolnitz' "disinterested attention," namely, "attending with no ulterior purpose."<sup>26</sup> Thus, the question to ask is "How does one attend to (read) a poem or any

<sup>20</sup> "Contextualism Reconsidered," *The Journal of Aesthetics and Art Criticism*, vol. 18 (1959), pp. 224–225.

<sup>21</sup> *Ibid.*, p. 225.

<sup>22</sup> *Loc. cit.*

<sup>23</sup> *Ibid.*, p. 227.

<sup>24</sup> *Ibid.*, p. 237.

<sup>25</sup> *Ibid.*, p. 228. (Italics mine.)

<sup>26</sup> Vivas' remark about the improbability of being able to read *The Brothers Karamazov* as art suggests that "intransitive attention" may sometimes mean for him "that which can be attended to at one time" or "that which can be held before the mind at one time." However, this second possible meaning is not one which is relevant here.

literary work transitively?" One can certainly attend to (read) a poem for a variety of different purposes and because of a variety of different reasons, but can one attend to a poem transitively? I do not think so, but let us consider the examples Vivas offers. He mentions "a type of reader" who uses a poem or parts of a poem as a spring-board for "loose, uncontrolled, relaxed day-dreaming, wool-gathering rambles, free from the contextual control" of the poem.<sup>27</sup> But surely it would be wrong to say such musing is a case of transitively attending to a poem, since it is clearly a case of not attending to a poem. Another supposed way of attending to a poem transitively is by approaching it "as diagnostic evidence of the author's neuroses." Vivas is right if he means that there is no critical point in doing this since it does not throw light on the poem. But this is a case of *using* information gleaned from a poem to make inferences about its author rather than attending to a poem. If anything can be said to be attended to here it is the author's neuroses (at least they are being thought about). This kind of case is perhaps best thought of as a rather special way of getting distracted from a poem. Of course, such "biographical" distractions might be insignificant and momentary enough so as scarcely to distract attention from the poem (a flash of insight or understanding about the poet). On the other hand, such distractions may turn into dissertations and whole careers. Such an interest may lead a reader to concentrate his attention (when he does read a poem) on certain "informational" aspects of a poem and to ignore the remaining aspects. As deplorable as such a sustained practice may be, it is at best a case of attending to certain features of a poem and ignoring others.

Another way that poetry may allegedly be read transitively is by reading it as history. This case is different from the two preceding ones since poetry often *contains* history (makes historical statements or at least references) but does not (usually) contain statements about the author's neuroses and so on nor does it contain statements about what a reader's free associations are about (otherwise we would not call them "*free associations*"). Reading a poem as history suggests that we are attending to (thinking about) historical events by way of attending to a poem—the poem is a time-telescope. Consider the following two sets of lines:

In fourteen hundred and ninety-two  
Columbus sailed the ocean blue.

<sup>27</sup> Vivas, op. cit., p. 231.

Or like stout Cortez when with eagle eyes  
He star'd at the Pacific—and all his men  
Look'd at each other with a wild surmise—  
Silent, upon a peak in Darien.

Someone might read both of these raptly and not know that they make historical references (inaccurately in one case)—might this be a case of intransitive attention? How would the above reading differ—so far as attention is concerned—from the case of a reader who recognized the historical content of the poetic lines? The two readings do not differ as far as attention is concerned. History is a part of these sets of poetic lines and the two readings differ in that the first fails to take account of an aspect of the poetic lines (its historical content) and the second does not fail to do so. Perhaps by "reading as history" Vivas means "reading *simply* as history." But even this meaning does not mark out a special kind of attention but rather means that only a single aspect of a poem is being noticed and that its rhyme, meter, and so on are ignored. Reading a poem as social criticism can be analyzed in a fashion similar to reading as history. Some poems simply are or contain social criticism, and a complete reading must not fail to notice this fact.

The above cases of alleged interested attending can be sorted out in the following way. Jones listening to the music and our playwright watching the rehearsal are both attending with ulterior motives to a work of art, but there is no reason to suppose that the attention of either is different in kind from that of an ordinary spectator. The reader who reads a poem as history is simply attending to an aspect of a poem. On the other hand, the remaining cases—Jones beside the painting telling of his grandfather, the gloating impresario, daydreaming while "reading" a poem, and so on—are simply cases of not attending to the work of art.

In general, I conclude that "disinterestedness" or "intransitiveness" cannot properly be used to refer to a special kind of attention. "Disinterestedness" is a term which is used to make clear that an action has certain kinds of motives. Hence, we speak of disinterested findings (of boards of inquiry), disinterested verdicts (of judges and juries), and so on. Attending to an object, of course, has its motives but the attending itself is not interested or disinterested according to whether its motives are of the kind which motivate interested or disinterested action (as findings and verdicts might), although the attending may be more or less close.

I have argued that the second way of conceiving the aesthetic attitude is also a myth, or at least that its main content—disinterested attention—is; but I must now try to establish that the view misleads aesthetic theory. I shall argue that the attitude-theorist is incorrect about (1) the way in which he wishes to set the limits of aesthetic relevance; (2) the relation of the critic to a work of art; and (3) the relation of morality to aesthetic value.

Since I shall make use of the treatment of aesthetic relevance in Jerome Stolnitz' book, let me make clear that I am not necessarily denying the relevance of the specific items he cites but disagreeing with his criterion of relevance. His criterion of relevance is derived from his definition of "aesthetic attitude" and is set forth at the very beginning of his book. This procedure leads Monroe Beardsley in his review of the book to remark that Stolnitz' discussion is premature.<sup>28</sup> Beardsley suggests "that relevance cannot be satisfactorily discussed until after a careful treatment of the several arts, their dimensions and capacities."<sup>29</sup>

First, what is meant by "aesthetic relevance"? Stolnitz defines the problem by asking the question: "Is it ever 'relevant' to the aesthetic experience to have thoughts or images or bits of knowledge which are not present within the object itself?"<sup>30</sup> Stolnitz begins by summarizing Bullough's experiment and discussion of single colors and associations.<sup>31</sup> Some associations absorb the spectator's attention and distract him from the color and some associations "fuse" with the color. Associations of the latter kind are aesthetic and the former are not. Stolnitz draws the following conclusion about associations:

If the aesthetic experience is as we have described it, then whether an association is aesthetic depends on whether it is compatible with the attitude of "disinterested attention." If the association re-enforces the focusing of attention upon the object, by "fusing" with the object and thereby giving it added "life and significance," it is genuinely aesthetic. If, however, it arrogates attention to itself and away from the object, it undermines the aesthetic attitude.<sup>32</sup>

It is not clear how something could *fuse* with a single color, but "fusion" is one of those words in aesthetics which is rarely defined. Stolnitz then makes use of a more fruitful example, one from I. A. Richards' *Practical Criticism*.<sup>33</sup> He cites the responses of students to the poem which begins:

Between the erect and solemn trees  
I will go down upon my knees;  
I shall not find this day  
So meet a place to pray.

The image of a rugby forward running arose in the mind of one student-reader on reading the third verse of this poem. A cathedral was suggested to a second reader of the poem. The cathedral image "is congruous with both the verbal meaning of the poem and the emotions and mood which it expresses. It does not divert attention away from the poem."<sup>34</sup> The rugby image is presumably incongruous and diverts attention from the poem.

It is a confusion to take compatibility with disinterested attention as a criterion of relevance. If, as I have tried to show, *disinterested attention* is a confused notion, then it will not do as a satisfactory criterion. Also, when Stolnitz comes to show why the cathedral image is, and the rugby image is not relevant, the criterion he actually uses is *congruousness with the meaning of the poem*, which is quite independent of the notion of disinterestedness. The problem is perhaps best described as the problem of relevance to a poem, or more generally, to a work of art, rather than aesthetic relevance.

A second way in which the attitude theory misleads aesthetics is its contention that a critic's relationship to a work of art is different in kind from the relationship of other persons to the work. H. S. Langfeld in an early statement of this view wrote that we may "slip from the attitude of aesthetic enjoyment to the attitude of the critic." He characterizes the critical attitude as "intellectually occupied in coldly estimating . . . merits" and the aesthetic attitude as responding "emotionally to" a work of art.<sup>35</sup> At the beginning of his book in the discussion of the aesthetic attitude, Stolnitz declares that if a percipient of a work of art "has the purpose of passing judgment upon it, his attitude is not aesthetic."<sup>36</sup> He develops this line at a later stage of his book, arguing that appreciation (perceiving with the aesthetic attitude) and criticism (seeking for reasons to support an evaluation of a work) are (1) distinct and (2) "psychologically opposed to each other."<sup>37</sup> The critical attitude is questioning, analytical, probing for strengths and weakness, and so on. The aesthetic attitude is just the opposite: "It commits our allegiance to the object freely and unquestioningly"; "the spectator 'surrenders' himself to the work of art."<sup>38</sup> "Just because the two

<sup>28</sup> *The Journal of Philosophy*, vol. 57 (1960), p. 624.

<sup>29</sup> *Ibid.*, pp. 54-55.

<sup>30</sup> *Op. cit.*, p. 35.

<sup>31</sup> *Ibid.*, pp. 55-56.

<sup>32</sup> *Ibid.*, p. 377.

<sup>33</sup> *Loc. cit.*

<sup>34</sup> *Ibid.*, p. 56.

<sup>35</sup> *Ibid.*, pp. 377-378.

<sup>36</sup> *Op. cit.*, p. 53.

<sup>37</sup> *Ibid.*, p. 54.

<sup>38</sup> *The Aesthetic Attitude* (New York, 1920), p. 79.

attitudes are inimical, whenever criticism obtrudes, it reduces aesthetic interest."<sup>39</sup> Stolnitz does not, of course, argue that criticism is unimportant for appreciation. He maintains criticism plays an important and necessary role in preparing a person to appreciate the nuances, detail, form, and so on of works of art. We are quite right, he says, thus to read and listen perceptively and acutely, but he questions, "Does this mean that we must analyze, measure in terms of value-criteria, etc., *during* the supposedly aesthetic experience?"<sup>40</sup> His answer is "No" and he maintains that criticism must occur "*prior* to the aesthetic encounter,"<sup>41</sup> or it will interfere with appreciation.

How does Stolnitz know that criticism will always interfere with appreciation? His conclusion sounds like one based upon the observations of actual cases, but I do not think it is. I believe it is a logical consequence of his definition of aesthetic attitude in terms of disinterested attention (no ulterior purpose). According to his view, to appreciate an object aesthetically one has to perceive it with no ulterior purpose. But the critic has an ulterior purpose—to analyze and evaluate the object he perceives—hence, in so far as a person functions as a critic he cannot function as an appreciator. But here, as previously, Stolnitz confuses a perceptual distinction with a motivational one. If it were possible to *attend* disinterestedly or interestedly, then perhaps the critic (as percipient) would differ from other percipients. But if my earlier argument about attending is correct, the critic differs from other percipients only in his motives and intentions and not in the way in which he attends to a work of art.

Of course, it might just be a fact that the search for reasons is incompatible with the appreciation of art, but I do not think it is. Several years ago I participated in a series of panel discussions of films. During the showing of each film we were to discuss, I had to take note of various aspects of the film (actor's performance, dramatic development, organization of the screen-plane and screen-space at given moments, and so on) in order later to discuss the films. I believe that this practice not only helped educate me to appreciate subsequent films but that it enhanced the appreciation of the films I was analyzing. I noticed and was able to appreciate things about the films I was watching which ordinarily out of laziness I would not have noticed. I see no reason why the same should not be the case with the professional critic or any critical

percipient. If many professional critics seem to appreciate so few works, it is not because they are critics, but perhaps because the percentage of good works of art is fairly small and they suffer from a kind of combat fatigue.

I am unable to see any significant difference between "perceptively and acutely" attending to a work of art (which Stolnitz holds enhances appreciation) and searching for reasons, so far as the experience of a work of art is concerned. If I attend perceptively and acutely, I will have certain standards and/or paradigms in mind (not necessarily consciously) and will be keenly aware of the elements and relations in the work and will evaluate them to some degree. Stolnitz writes as if criticism takes place and then is over and done with, but the search for and finding of reasons (noticing this fits in with that, and so on) is continuous in practiced appreciators. A practiced viewer does not even have to be looking for a reason, he may just notice a line or an area in a painting, for example, and the line or area becomes a reason why he thinks the painting better or worse. A person may be a critic (not necessarily a good one) without meaning to be or without even realizing it.

There is one final line worth pursuing. Stolnitz' remarks suggest that one reason he thinks criticism and appreciation incompatible is that they compete with one another for time (this would be especially bad in the cases of performed works). But seeking and finding reasons (criticism) does not compete for time with appreciation. First, to seek for a reason means to be ready and able to notice something and to be thus ready and able as one attends does not compete for time with the attending. In fact, I should suppose that seeking for reasons would tend to focus attention more securely on the work of art. Second, finding a reason is an achievement, like winning a race. (It takes time to run a race but not to win it.) Consider the finding of the following reasons. How much time does it take to "see" that a note is off key (or on key)? How long does it take to notice that an actor mispronounces a word (or does it right)? How much time does it take to realize that a character's action does not fit his already established personality? (One is struck by it.) How long does it take to apprehend that a happy ending is out of place? It does not take time to find any of these reasons or reasons in general. Finding a reason is like coming to understand—it is done in a flash. I do not mean to suggest that one cannot be mistaken in finding a

<sup>39</sup> Ibid., p. 379.

<sup>40</sup> Ibid., p. 380.

<sup>41</sup> Loc. cit.

reason. What may appear to be a fault or a merit (a found reason) in the middle of a performance (or during one look at a painting and so forth) may turn out to be just the opposite when seen from the perspective of the whole performance (or other looks at the painting).

A third way in which the attitude theory misleads aesthetic theory is its contention that aesthetic value is always independent of morality. This view is perhaps not peculiar to the attitude theory, but it is a logical consequence of the attitude approach. Two quotations from attitude-theorists will establish the drift of their view of morality and aesthetic value.

We are either concerned with the beauty of the object or with some other value of the same. Just as soon, for example, as ethical considerations occur to our mind, our attitude shifts.<sup>43</sup>

Any of us might reject a novel because it seems to conflict with our moral beliefs . . . When we do so . . . We have *not* read the book aesthetically, for we have interposed moral . . . responses of our own which are alien to it. This disrupts the aesthetic attitude. We cannot then say that the novel is *aesthetically* bad, for we have not permitted ourselves to consider it aesthetically. To maintain the aesthetic attitude, we must follow the lead of the object and respond in concert with it.<sup>43</sup>

This conception of the aesthetic attitude functions to hold the moral aspects and the *aesthetic* aspects of the work of art firmly apart. Presumably, although it is difficult to see one's way clearly here, the moral aspects of a work of art cannot be an object of aesthetic attention because aesthetic attention is by definition disinterested and the moral aspects are somehow practical (interested). I suspect that there are a number of confusions involved in the assumption of the incompatibility of aesthetic attention and the moral aspects of art, but I shall not attempt to make these clear, since the root of the assumption—disinterested attention—is a confused notion. Some way other than in terms of the aesthetic attitude, then, is needed to discuss the relation of morality and aesthetic value.

David Pole in a recent article<sup>44</sup> has argued that the moral vision which a work of art may embody is *aesthetically* significant. It should perhaps be remarked at this point that not all works of art embody a moral vision and perhaps some kinds of art (music, for example) cannot embody a moral vision, but certainly some novels, some poems, and

some films and plays do. I assume it is unnecessary to show how novels and so on have this moral aspect. Pole notes the curious fact that while so many critics approach works of art in "overtly moralistic terms," it is a "philosophical commonplace . . . that the ethical and the aesthetic modes . . . form different categories."<sup>45</sup> I suspect that many philosophers would simply say that these critics are confused about their roles. But Pole assumes that philosophical theory "should take notice of practice"<sup>46</sup> and surely he is right. In agreeing with Pole's assumption I should like to reserve the right to argue in specific cases that a critic may be misguided. This right is especially necessary in a field such as aesthetics because the language and practice of critics is so often burdened with ancient theory. Perhaps *all* moralistic criticism is wrong but philosophers should not rule it out of order at the very beginning by use of a definition.

Pole thinks that the moral vision presented by a particular work of art will be either true or false (perhaps a mixture of true and false might occur). If a work has a false moral vision, then something "is lacking within the work itself. But to say that is to say that the [work] is internally incoherent; some particular aspect must jar with what—on the strength of the rest—we claim a right to demand. And here the moral fault that we have found will count as an aesthetic fault too."<sup>47</sup> Pole is trying to show that the assessment of the moral vision of a work of art is just a special case of coherence or incoherence, and since everyone would agree that coherence is an aesthetic category, the assessment of the moral vision is an aesthetic assessment.

I think Pole's conclusion is correct but take exception to some of his arguments. First, I am uncertain whether it is proper to speak of a moral vision being true or false, and would want to make a more modest claim—that a moral vision can be judged to be acceptable or unacceptable. (I am not claiming Pole is wrong and my claim is not inconsistent with his.) Second, I do not see that a false (or unacceptable) moral vision makes a work incoherent. I should suppose that to say a work is coherent or incoherent is to speak about how its parts fit together and this involves no reference to something outside the work as the work's truth or falsity does.

In any event, it seems to me that a faulty moral

<sup>43</sup> H. S. Langfeld, *op. cit.*, p. 73.

<sup>44</sup> J. Stolnitz, *op. cit.*, p. 36.

<sup>45</sup> "Morality and the Assessment of Literature," *Philosophy*, vol. 37 (1962), pp. 193–207.

<sup>46</sup> *Ibid.*, p. 193.

<sup>47</sup> *Loc. cit.*

<sup>48</sup> *Ibid.*, p. 206.

vision can be shown to be an aesthetic fault independently of Pole's consideration of truth and coherence. As Pole's argument implies, a work's moral vision is a *part* of the work. Thus, any statement—descriptive or evaluative—about the work's moral vision is a statement about the *work*; and any statement about a *work* is a critical statement and, hence, falls within the aesthetic domain. To judge a moral vision to be morally unacceptable is to judge it defective and this amounts to saying that the work of art has a defective part. (Of course, a judgment of the acceptability of a moral vision may be wrong, as a judgment of an action sometimes is, but this fallibility does not make any difference.) Thus, a work's moral vision may be an aesthetic merit or defect just as a work's degree of unity is a merit or defect. But what justifies saying that a moral vision is a part of a work of art? Perhaps "part" is not quite the right word but it serves to make the point clear enough. A novel's moral vision is an essential part of the novel and if it were removed (I am not sure how such surgery could be carried out) the novel would be greatly changed. Anyway, a novel's moral vision is not like its covers or binding. However, someone might still argue that even though a work's moral vision is defective and the moral vision is part of the work, that this defect is not an *aesthetic* defect. How is "aesthetic" being used here? It is being used to segregate certain aspects or parts of works of art such as formal and stylistic aspects from such aspects as a work's moral vision. But it seems to me that the separation is only nominal. "Aesthetic" has been selected as a name for a certain sub-set of characteristics of works of art. I certainly cannot object to such a stipulation, since an underlying aim of this essay is to suggest the vacuousness of the term "aesthetic." My concern at this point is simply to insist that a work's moral vision is a part of the work and that, therefore, a critic can legitimately describe and evaluate it. I would *call* any defect or merit which a critic can legitimately point out an aesthetic defect or merit, but what we call it does not matter.

It would, of course, be a mistake to judge a work solely on the basis of its moral vision (it is only one part). The fact that some critics have judged works of art in this way is perhaps as much responsible as the theory of aesthetic attitude for the attempts to separate morality from the aesthetic. In fact, such criticism is no doubt at least partly responsible for the rise of the notion of the aesthetic attitude.

<sup>48</sup> Tomas, *op. cit.*, p. 63.

If the foregoing arguments are correct, the second way of conceiving the aesthetic attitude misleads aesthetic theory in at least three ways.

### III

In answer to a hypothetical question about what is seen in viewing a portrait with the aesthetic attitude, Tomas in part responds "If looking at a picture and attending closely to how it looks is not really to be in the aesthetic attitude, then what on earth is?"<sup>48</sup> I shall take this sentence as formulating the weakest version of the aesthetic attitude. (I am ignoring Tomas' distinction between appearance and reality. See footnote 7. My remarks, thus, are not a critique of Tomas' argument; I am simply using one of his sentences.) First, this sentence speaks only of "looking at a picture," but "listening to a piece of music," "watching and listening to a play," and so on could be added easily enough. After thus expanding the sentence, it can be contracted into the general form: "Being in the aesthetic attitude is attending closely to a work of art (or a natural object)."

But the aesthetic attitude ("the hallmark of modern aesthetics") in this formulation is a great letdown—it no longer seems to say anything significant. Nevertheless, this does seem to be all that is left after the aesthetic attitude has been purged of *distancing* and *disinterestedness*. The only thing which prevents the aesthetic attitude from collapsing into simple attention is the qualification *closely*. One may, I suppose, attend to a work of art more or less closely, but this fact does not seem to signify anything very important. When "being in the aesthetic attitude" is equated with "attending (closely)," the equation neither involves any mythical element nor could it possibly mislead aesthetic theory. But if the definition has no vices, it seems to have no virtues either. When the aesthetic attitude finally turns out to be simply attending (closely), the final version should perhaps not be called "the weakest" but rather "the vacuous version" of the aesthetic attitude.

Stolnitz is no doubt historically correct that the notion of the aesthetic attitude has played an important role in the freeing of aesthetic theory from an overweening concern with beauty. It is easy to see how the slogan, "Anything can become an object of the aesthetic attitude," could help accomplish this liberation. It is worth noting, however, that the same goal could have been (and

perhaps to some extent was) realized by simply noting that works of art are often ugly or contain ugliness, or have features which are difficult to include within beauty. No doubt, in more recent times people have been encouraged *to take an aesthetic attitude toward a painting* as a way of lowering

their prejudices, say, against abstract and non-objective art. So if the notion of aesthetic attitude has turned out to have no theoretical value for aesthetics, it has had practical value for the appreciation of art in a way similar to that of Clive Bell's suspect notion of significant form.

*Washington State University*

## VI. THE UNEXPECTED EXAMINATION

BRIAN MEDLIN

A SCHOOLMASTER tells his class that he is going to give them an examination on one of the five days of the following week. (For our convenience later, let it be an afternoon examination.) He adds that it will be unexpected. By this he means that at no stage before the examination is actually given will it be possible to determine the day on which it will be given. The boys argue: "If the examination is held on the last day, then on the morning of the last day it will be possible to determine when it will be held. Hence, if the examination is to be unexpected, it cannot be given on the last day. This leaves only the first four days as possibilities. By the same reasoning, the examination cannot be given on the fourth day. And so on to the third, second, and first days in turn. On no day out of five, then, is it possible to give an unexpected examination." As it happens, the examination is given on the second day. Perhaps the students are taken by surprise, perhaps not; this is not important. What is important is that if the examination is given on the second day, then it is unexpected in the required sense. It is not possible to determine beforehand when the examination will be held. Hence, in spite of the boys' argument, what the master undertakes to do is possible. He does it. What has gone wrong?

Several philosophers have discussed this problem in *Mind*.<sup>1</sup> Of these, we must put Mr. Shaw first and the rest nowhere. Shaw has at least got onto the right track. His successors cross that track at right angles. This paper works out the solution along lines drawn in Shaw's article.

<sup>1</sup> D. J. O'Connor, "Pragmatic Paradoxes," 1948, pp. 358-359; "Pragmatic Paradoxes and Fugitive Propositions," 1951, pp. 536-538; L. J. Cohen, "Mr. O'Connor's 'Pragmatic Paradoxes,'" 1950, pp. 85-87; Peter Alexander, "Pragmatic Paradoxes," 1950, pp. 536-538; M. Scriven, "Paradoxical Announcements," 1951, pp. 403-407; Paul Weiss, "The Prediction Paradox," 1952, pp. 265-269; W. V. Quine, "On a So-called Paradox," 1953, pp. 65-67; R. Shaw, "The Paradox of the Unexpected Examination," 1958, pp. 382-384; Ardon Lyon, "The Prediction Paradox," 1959, pp. 510-517; G. C. Nerlich, "Unexpected Examinations and Unprovable Statements," 1961, pp. 503-513. Hereafter, all references by year and page will be to *Mind*.

Too many of these authors show that they have not the least idea of how to go about solving a paradox and presenting the solution. The rules they follow seem to be these:—

- (i) State the paradox.
- (ii) Say that (you like to think that) the trouble is so-and-so.
- (iii) Chat vaguely about so-and-so.

Above all,

- (iv) Avoid any attempt to restate the paradoxical argument in the terms of your proposed solution showing that the result is either an invalid argument or a valid argument to a harmless conclusion.

And even above (iv),

- (v) Refrain from wondering whether the paradox might not be generalized in such a way that it resists your proposed solution.

It may be that there are two paradoxes, not one; that Quine (1953) really does solve one of them; and that this one is, as he says, a trivial paradox. If so, then some of my remarks in this introduction are unfair to Quine and perhaps also to others. But I expect his reputation, at any rate, to survive them. And I am not very worried about being unfair to someone who picks out a trivial paradox for solution, leaving the more difficult one unsolved and unnoticed. I am less worried still in the case of those who muffle the solution of even the simple paradox.

The paradox may or may not be an interesting one. But this is something that cannot be known in advance of a solution. Those philosophers who find it boring have presumably (unlike Quine) the right answer. Be that as it may, the right answer is not so easy to guess as some people imagine. In conversation, I have not found anyone who started off by poo-h-pooing the paradox as a childish puzzle able to solve it. Most attempts have involved mistakes in elementary logic. Surprisingly, the commonest of these has been to suppose that what follows from a conditional proposition must depend on the antecedent of the conditional as assumption. When I use the premiss:

If the examination does not occur on one of the first four days, then on the morning of the fifth day it will be possible to say that it will occur on that day.

and the premiss:

It is not possible to determine the day on which the examination will occur at any time before it does occur.

and *no other* premiss, then I use *those* premisses (and no other). I do not at the same time use the premiss:

The examination does not occur on any of the first four days.

This means that I do not have to assume that the examination must occur on the last day to prove that it cannot. (Even if I did, this would be very far from a solution of the paradox.) Some readers will be impatient with this point and those who are not ought to be. But I have had to make it so often in conversation that I cannot doubt the need for it here.

In fact, the solution is a little disappointing in that it shows that we have nothing to learn from the paradox that was not known before. Certainly, we do not have to learn what Mr. Nerlich<sup>2</sup> would teach us (1961: 513), that there are contradictions which are true after all and true *because* they are contradictions:

In the case we have been considering all along, one person is the source of all we know, and when our only source of information seems to impeach itself, then we do not know what to make of it. That is precisely what this particular source of knowledge wants to achieve. One way of saying nothing is to contradict yourself. And if you manage to contradict yourself by saying that you are saying nothing, then you do not, in the end, contradict yourself at all, you *can* eat your cake and have it too.

This solution is surely just a reformulation of the paradox. If we finish up thus where we began, then it is clear that we have not eliminated the fallacy from our original argument. I shall refer to Nerlich's article throughout this paper. It is convenient to make the point that the paradox is essentially one of self-reference by exposing his mistakes (or what I take to be his mistakes). I shall do this in Section I.

In the second Section, I produce a generalization of the paradox. The primary purpose here is again to defend my account of it.<sup>3</sup> This could be done without going beyond finite order types. The generalization for infinite sets is undertaken purely for the fun of it.

In Section I, I shall reduce the number of days to 3. It could be reduced to 2, 1, or even to 0. I choose 3 largely for polemical reasons.

I shall use the following logical symbols:

"—"	(standing above a symbol) for propositional negation
"∨"	for non-exclusive disjunction
"&"	for conjunction
"⊃"	for material implication
"(E $v$ )"	(where $v$ is some variable) for existential quantification
"(E! $v$ )"	(where $v$ is some variable) for existential quantification with uniqueness
"(U $v$ )"	(where $v$ is some variable) for universal quantification
"ε"	for class membership
"⊂"	for class inclusion

In addition, I use "⊢" in expressions of the form " $A_1, \dots, A_n \vdash B$ " to mean " $B$  is deducible from one or more of  $A_1, \dots, A_n$  as premisses."

## I

Denote the days on which the examination may occur by  $d_1, d_2, d_3$ . Let  $p_i$  be the proposition: *The examination occurs on  $d_i$* , ( $1 \leq i \leq 3$ ). Let the day on which the examination occurs be  $d_x$ . Surely we must understand unexpectedness in such a way that from " $p_x$  is unexpected" it follows that  $\bar{p}_3$ . For

$$p_1 \vee p_2 \vee p_3, \bar{p}_1 \vee \bar{p}_2 \vdash p_3$$

This means that if  $p_3$ , then on the morning of  $d_3$  we should be able to conclude that  $p_3$ , contrary to the condition that  $p_x$  is unexpected. Further, from  $\bar{p}_1$  we can derive  $p_2$ . For we already have  $\bar{p}_3$  and

$$p_1 \vee p_2 \vee p_3, \bar{p}_1, \bar{p}_3 \vdash p_2$$

This means that if  $p_2$ , then on the morning of  $d_2$  we may conclude that  $p_2$ , contrary to the condition that  $p_x$  is unexpected. It follows that  $\bar{p}_2$ . In this way we may conclude also that  $\bar{p}_1$ .

But must we understand unexpectedness in such a way as to produce this result? We are told that at no stage before the examination will it be possible to determine  $x$ . We take this not as a remark about our mental powers, but as

(M) *The information concerning  $d_x$  is not sufficient to allow the determination of  $x$  at any stage before the examination is actually given.*

<sup>2</sup> I have corresponded with Nerlich about this paradox. (In fact it was correspondence with Nerlich that first got me interested in it.) He has read a copy of this paper and generously seconds my remarks on his article.

<sup>3</sup> More accurately, perhaps: Shaw's account of it.

This meta-statement is true provided only that  $x \neq 3$ . For the information<sup>4</sup> concerning  $d_x$  is

(I)  $(p_1 \vee p_2 \vee p_3) \& \overline{p_i} \& \overline{p_j} \ (i \neq j; 1 \leq i, j \leq 3)$

For the case  $x = 3$ , (M) is false since

$$p_1 \vee p_2 \vee p_3, \overline{p_1} \vee \overline{p_2} \vdash p_3$$

If the paradox is to arise now, the boys' argument must show that (M) is false for every possible case and it must show this in the teeth of the fact that (M) is true for two out of the three possible cases. Let us see how this might be done.

It is obvious that

$$p_1 \vee p_2 \vee p_3, \overline{p_1} \vee \overline{p_2} \vdash p_3$$

Hence it follows that if  $p_3$ , then on the morning of  $d_3$  we might conclude that  $p_3$ , i.e., that  $x = 3$ . But this contradicts (M). It follows that  $\overline{p_3}$ . It is also obvious that

$$p_1 \vee p_2 \vee p_3, \overline{p_1}, \overline{p_3} \vdash p_2$$

We already have the result that  $\overline{p_3}$ . Hence it follows that if  $p_2$ , then on the morning of  $d_2$ , we might conclude that  $p_2$ . But this contradicts (M). It follows that  $\overline{p_2}$ . In the same way, we may conclude also that  $\overline{p_1}$ .

There is something wrong with this argument. Notice that

$$p_1 \vee p_2 \vee p_3, \overline{p_1} \vdash p_2$$

is false. To deduce  $p_2$  we need as additional premiss  $\overline{p_3}$  which depends in turn on (M). The argument employs (M) as though it were itself part of the information referred to by (M). For if (M) were additional to that information, there would be no contradiction of (M) involved in

$$p_1 \vee p_2 \vee p_3, \overline{p_1}, (M) \vdash p_2$$

together with  $p_2$ . It is not inconsistent with " $p_2$  cannot be deduced from the information:  $p_1 \vee p_2 \vee p_3, \overline{p_1}$ " to say that " $p_2$  can be deduced from the information:  $p_1 \vee p_2 \vee p_3, \overline{p_1}, (M)$ ."

The first of Nerlich's two objections to Shaw is that "he seems to take it for granted that no statement may refer to itself, though Goedel would seem to have demonstrated that at least one can"

(1961: 105). It does not seem to me that Goedel has demonstrated this. I would say, rather, that a great part of Goedel's argument consists in the construction of machinery to eliminate the kind of self-reference we have here. I agree with Nerlich that Shaw is not entitled to take it for granted, as he does, that no statement can be self-referring. It all depends on what you mean by "self-referring." (See, for instance, R. M. Smullyan, "Languages in which Self-reference is Possible," *The Journal of Symbolic Logic*, 1957, pp. 55-67.) Nonetheless, even granted perfectly proper cases of self-reference, it is still open to us to show that particular cases of self-reference are improper. The present case of self-reference is improper.

The proposition (M) says something about the propositions in a non-empty set  $S$ ; namely, that the conjunction of all these propositions does not constitute a premiss of sufficient power to permit the determination of  $x$  at any stage before the examination is given. (Where  $x = 1$ , the words "at any stage before the examination is given" have little point. Otherwise, they may be filled in according to the formula "even in conjunction with the additional premiss  $p_1 \vee \dots \vee p_i$ , where  $1 \leq i \leq x$ ." But if (M) is in  $S$ , then what (M) says is (roughly) that (M) does not permit us to determine  $x$ . This kind of self-reference is circular. It invites the question, "What does not permit us to determine  $x$ ?" We do not understand (M) if we cannot answer this question. But the answer is "(M)." Alternatively, we may say that we do not understand (M) until we know what (M) is about, which set  $S$  happens to be. If (M) is itself in  $S$ , then we shall never know this and never understand (M).

Nerlich's second, and more serious, objection is that "the self-referential feature is quite inessential to the paradox." How much is in this? Suppose that the boys develop the following argument. We have been given the information (I).

<sup>4</sup> The uniqueness condition (second conjunct) in (I) is not essential. If it did not serve conveniently to introduce this footnote, it could be dropped. It has been pointed out by the referee that, for any natural numbers  $n$  and  $m$ , the schoolmaster could announce  $n$  unexpected examinations to be given in the next  $m$  days (no two on the same day). In all cases where  $m < n$  the solution of the paradox would be trivial. But for cases where  $m > n$  the paradoxical argument would commence as follows: "Consider the first examination. Let it occur on  $d_m$ . Clearly, since  $p_2$  is unexpected, we know that  $x \neq m - n + 1$ ; for otherwise it would be possible on the morning of  $d_{m-n-1}$  to predict that the examination would occur on that day." Then the argument would develop as I have developed it in the text.

Similarly, in Section II where I generalize the paradox for infinite sets we could drop the uniqueness requirement. Here we could work with  $n$  elements of  $S$  where  $n$  is any natural number. We would consider  $\sigma$ , the  $R$ -first element of  $S$  to have  $P$ . We would consider also the set:

$$S = S - \{s_1, s_2, \dots, s_{n-1}\}$$

where  $s_1, s_2, \dots, s_{n-1}$  are the first  $n-1$  elements of the  $SR$ -series defined on p. 70. Then we would reason concerning  $\sigma$  and  $S$  as I have reasoned concerning  $s$  and  $S$ . I have not generalized the paradox in this way in the actual text for the reason that to do so would require some (slight) complication of the argument. The new generalization is irrelevant to the solutions I want to dispose of.

We have also been told that  $p_x$  is unexpected. Accordingly, to this information we may add

(M<sub>1</sub>) *From (I) it is not possible to determine  $x$ , even given as additional information one of*

$$\bar{p}_1, \bar{p}_1 \vee \bar{p}_2.$$

Since

$$(I), \bar{p}_1 \vee \bar{p}_2 \vdash \bar{p}_3$$

the proposition (M<sub>1</sub>) is false for the case  $x = 3$ . Hence

$$(1) \quad (M_1) \vdash \bar{p}_3$$

We may add further

(M<sub>2</sub>) *From (I) & (M<sub>1</sub>) it is not possible to determine  $x$ , even given as additional information*

$$\bar{p}_1.$$

Since from (1) and (I),  $\bar{p}_1, \bar{p}_3 \vdash \bar{p}_2$  we have (I),  $\bar{p}_1, (M_1) \vdash \bar{p}_2$ , the proposition (M<sub>2</sub>) is false for the case  $x = 2$ .

Hence

$$(2) \quad (M_2) \vdash \bar{p}_2$$

Finally, we may add

(M<sub>3</sub>) *From (I) & (M<sub>1</sub>) & (M<sub>2</sub>) it is not possible to determine  $x$ .*

From here the boys may continue to argue in the same way. From (1), (2), and

$$(I), \bar{p}_3, \bar{p}_2 \vdash \bar{p}_1$$

we have

$$(I), (M_1), (M_2) \vdash \bar{p}_1,$$

so that (M<sub>3</sub>) is false for the case  $x = 1$ . Hence

$$(3) \quad (M_3) \vdash \bar{p}_1$$

But now we can deduce a contradiction from what the master says. For from (1), (2), (3), we have

$$(4) \quad (M_1), (M_2), (M_3) \vdash \bar{p}_1 \vee \bar{p}_2 \vee \bar{p}_3.$$

We also have

$$(5) \quad (I) \vdash p_1 \vee p_2 \vee p_3$$

From (4), (5) we get

$$(6) \quad (I), (M_1), (M_2), (M_3) \vdash \frac{(p_1 \vee p_2 \vee p_3) \& \bar{p}_1 \vee \bar{p}_2 \vee \bar{p}_3}{\text{contradiction}}$$

This argument is valid. The boys interpret the master's warning " $p_x$  is unexpected" as

$$(C) \quad (M_1) \& (M_2) \& (M_3)$$

If their interpretation is correct, then what the master says does yield a contradiction. But there is no worry for us here. For (M<sub>3</sub>) is straight-forwardly false. Contrary to (M<sub>3</sub>) we have

$$(I), (M_1), (M_2) \vdash p_3$$

Hence no contradiction is forced upon us. We have

shown that no examination can be unexpected in the present sense. Though the students will be unable to arrive at  $x = 2$ , say, from (I), this does not provide a model for (C). It gives a model for (M<sub>1</sub>).

Nerlich would attempt at this stage to show that the paradox is still with us. I adapt his words (p. 505) to my text.<sup>5</sup>

From this set of rules—(I), (M<sub>1</sub>), (M<sub>2</sub>), (M<sub>3</sub>)—a contradiction follows by a precisely parallel series of steps as are used to derive a contradiction from (I) and (M). Shaw claims that there is no paradox here since no unexpected examination can in fact be set on any of  $d_1, \dots, d_3$ . But, in fact one can be set. Certainly, it is possible for an examination to occur during the week say on  $d_3$ . And it is quite plain that there is no way of deducing from (I), (M<sub>1</sub>), (M<sub>2</sub>), (M<sub>3</sub>) that  $d_3$  will be the day.

Here Nerlich is arguing that "the self-referential feature is quite inessential to the paradox." Yet in order to keep the paradox alive, he goes in for a fast bit of self-reference on his own account. He says that an examination on  $d_3$  will be unexpected. But in what sense unexpected? In the sense, according to him, that  $p_3$  cannot be deduced from (I) & (C). To clear up a small point first: we do have

$$(I) \& (C) \vdash p_3$$

since the premiss is inconsistent. This is not, however, contrary to Nerlich. He means that we cannot use (C) as a (true) premiss to determine  $x$ , where  $x = 2$ . Here he is right. But if this is to be taken as providing a model for (I) & (C), then we must interpret (C) as saying of *itself* that it does not, with (I), constitute sufficient information for the determination of  $x$ . The statement for which  $p_3$  does provide a model is

(M<sub>4</sub>) *The conjunction (I) & (C) does not constitute sufficient information for the determination of  $x$ .*

Unlike (C), the statement (M<sub>4</sub>) is true. It is true because (C) is false. Nerlich confuses (M<sub>4</sub>) with (C). He is then led to say that (C) is true because it is false. We should notice in passing that the case  $p_1$  provides a model for (M<sub>4</sub>). So does the case  $p_2$ : that is why Nerlich finds that even an examination on  $d_3$  is unexpected.

## II

The main purpose of this section is to show that the paradox arises solely from considerations concerning what premisses may establish which conclusions.<sup>6</sup> This is done by generalizing the paradox

<sup>5</sup> In the rest of this section I discuss Nerlich's criticism of Shaw in terms of my own text, silently adapting his words for this purpose.

<sup>6</sup> It has nothing to do with who says what to whom: It is not a pragmatic paradox (O'Connor, Cohen). It does not arise from supposing that an unexpected examination is a special kind of examination (Scriven). It has nothing to do with announcement

in such a way that its essential features become obvious.

Clearly, there is nothing sacred about the numbers 5 or 3. What holds for them will hold for any natural number  $n$ . (Where  $n = 1$  or  $n = 0$  the solution is trivial.)

Consider any non-empty set  $S$  simply ordered by a relation  $R$ . For  $s', s'' \in S$ , where  $s'$  precedes  $s''$  under  $R$  we say " $s' <_R s''$ ." We shall speak of the *R-first* and *R-last* element of  $S$  (or of a non-empty subset of  $S$ ) and also of the *immediate R-predecessor* of an element of  $S$ . The meaning of these expressions should be intuitively clear. We stipulate that one and only one element<sup>7</sup> of  $S$  has a certain property  $P$ . (We write " $Px$ " in the usual way.) Let this element be  $s$ . Let  $P$  (or  $P_s$ ) be *unexpected* in the sense that from the premisses

(i)  $(E!x) (x \in S \ \& \ Px)$

(ii)  $Ps' \vee Ps'' \vee \dots \vee Ps' \dots'$

we obtain  $(s', s'', \dots, s' \dots' <_R s)$

although it is not possible to derive the conclusion

(iii)  $P_s$

We shall allow the possibility that  $Ps' \vee Ps'' \vee \dots \vee Ps' \dots'$  should be an "infinite disjunction"—even an uncountably infinite one. To do so we replace the premiss (ii) with

(ii')  $(y) [(y \subset S \ \& \ (z) (z \in y \supset z <_R s)) \supset (Ex) (x \in y \ \& \ Px)]$

Allowing this possibility, then even for the general case, if  $P_s$  is unexpected, we cannot have  $s$  as the *R-last* element of  $S$ . For otherwise, where  $S_1 = S - \{s\}$ , we should have

$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_1 \ \& \ Px) \vdash P_s$

This violates the condition of unexpectedness for  $P_s$ .

To begin developing the paradox, then, we need an *R-last* member  $s_1$  of  $S$ . And to continue we need an *R-last* member  $s_2$  of  $S_1$ . We notice that  $s_2$  will be the immediate *R-predecessor* of  $s_1$ . Next we need  $s_3$  such that  $s_3$  is the immediate *R-predecessor* of  $s_2$ . As soon as we come to an element of  $S$  (other than the *R-first*) which has no immediate *R-predecessor*, we must break off the development. We may sup-

pose that if ever the development is to terminate in the paradoxical result, we must come eventually to the *R-first* element of  $S$ . If so, then there is only one kind of order-type for  $S$  which satisfies our requirements.  $S$  must be an ordered  $n$ -tuple for finite  $n$ . But we can go further than this.

Certainly the paradox may be developed for any finite order-type  $n$ . But, as will be shown, it can be generalized for an infinite set  $S$ , provided that  $S$  is countable. Suppose that  $S$  is an infinite set with  $s$  the one and only element having  $P$ . Let  $P_s$  be unexpected in the sense defined. Suppose also that  $S$  has an *R-last* element and that every element of  $S$  has an immediate *R-predecessor*.

Construct the series

$$S_R = \langle s_1, s_2, \dots, s_n, \dots \rangle$$

according to the rule

(a)  $s_1 = s$

(b)  $s_{n+1}$  is the immediate *R-predecessor* of  $s_n$ .

We can establish by induction on indices that no member of the  $S_R$ -series has  $P$ . Hence we can establish the paradoxical result that no member of  $S$  has  $P$ , provided only that every member of  $S$  comes into the  $S_R$ -series. This means that the paradox can be developed for any set of order-type  $^*\omega$ ; that is, for any set which is ordinally similar to the set of negative integers under the relation *less than*. Since any countably infinite set will receive the order-type  $^*\omega$  for a suitably chosen relation, the paradox may be developed for any such set. A condition that holds for each element of  $S$  will hold for each element of any subset of  $S$ . Hence if no element of  $S$  has  $P$ , no element of a finite  $n$ -membered subset of  $S$  has  $P$ . The paradoxical result for finite sets reappears as a consequence of the result for infinite sets. For given a set of  $n$  elements, we can always treat these as the last elements of a suitable set of order-type  $^*\omega$ .

It is easy to see how the paradox may be developed for ordered  $n$ -tuples. But here we are considering finite sets merely as subsets of infinite ones. Let us develop the paradox for the case where  $S$  is a set of order-type  $^*\omega$  under  $R$ .

(Scriven). Hence it has nothing to do with conditional declarations of intention (Alexander). It is concerned with order, but not necessarily with the temporal order: Hence it is not a paradox of prediction (Weiss, Lyon). Nor is it concerned with any real or supposed distinction between collective and distributive "or" (Weiss). No question of what so-and-so may know is relevant (Quine). Nor is any question about the source of our knowledge (Nerlich). All these results may be established at once, if we can produce a sufficiently abstract formulation of the paradox.

<sup>7</sup> This uniqueness condition could be dropped and (i) changed to

(i')  $(Ex) (x \in S \ \& \ Px)$

without blocking the argument to the paradoxical conclusion. I retain the condition here mainly so that I can conveniently refer the reader once again to footnote 4, p. 68. I retain the condition throughout the section because I retain it here.

We are told first that

$$(I^*) \quad (E!x) (x \in S \ \& \ Px)$$

Next we are told that

(M\*) *From the information given, it is not possible to determine which element of  $S$  has  $P$ , even in conjunction with information of the form*

$$(\overline{Ex}) (x \in S^* \ \& \ Px)$$

where  $S^*$  is a subset of  $S$  such that

$$(y) [(y \in S \ \& \ Py) \supset (z) (z \in S^* \supset z <_R y)]$$

The proposition (M\*) is a generalized version of (M). It is true for all possible cases except that where the  $R$ -last element of  $S$  has  $P$ .

Yet from (M\*) we may argue plausibly to the conclusion that no element of  $S$  has  $P$ . For let the  $R$ -last element of  $S$  be  $s_1$ . It is obvious that, for  $S_1 = S - \{s_1\}$

$$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_1 \ \& \ Px) \vdash Ps_1.$$

But if  $Ps_1$ , this violates (M\*). Hence  $\overline{Ps}_1$ .

Suppose now that we have established  $\overline{Ps}_1, \overline{Ps}_2, \dots, \overline{Ps}_n$ , where  $s_{i+1}$  is the immediate  $R$ -predecessor of  $s_i$  ( $1 \leq i < n$ ). Consider now  $s_{n+1}$ , the immediate  $R$ -predecessor of  $s_n$ . It is obvious that for  $S_{n+1} = S - \{s_1, \dots, s_{n+1}\}$

$$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_{n+1} \ \& \ Px) \vdash Ps_1 \vee \dots \vee Ps_n \vee Ps_{n+1}.$$

By the hypothesis of induction,  $\overline{Ps}_1 \vee \overline{Ps}_2 \vee \dots \vee \overline{Ps}_n$ . Hence, from the information given and the condition  $(\overline{Ex}) (x \in S_{n+1} \ \& \ Px)$  it follows that  $Ps_{n+1}$ . But if  $Ps_{n+1}$ , this result contradicts (M\*). Hence  $\overline{Ps}_{n+1}$ .

By the principle of mathematical induction, we would seem now to have established upon (M\*) that no element of  $S$  has  $P$ .

The fallacy in this argument occurs during the inductive step. Having established  $\overline{Ps}_1, \overline{Ps}_2, \dots, \overline{Ps}_n$ , we cannot establish  $\overline{Ps}_{n+1}$  by the method we have employed. This means that we cannot establish even  $\overline{Ps}_2$ . For consider the step where  $n = 1$ . It is taken in the following way. "It is obvious that for  $S_2 = S_1 - \{s_2\}$ ,

$$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_2 \ \& \ Px) \vdash Ps_1 \vee Ps_2$$

But we have already established that  $\overline{Ps}_1$ . So that from the information given and  $(\overline{Ex}) (x \in S_2 \ \& \ Px)$  we may derive  $Ps_2$ . But if  $Ps_2$ , this result contradicts (M\*). Hence  $\overline{Ps}_2$ ."

There is here only one explicit appeal to (M\*). Yet the conclusion depends on  $\overline{Ps}_1$  which depends in turn upon (M\*). For it is not true that

$$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_2 \ \& \ Px) \vdash Ps_2,$$

only that

$$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_2 \ \& \ Px), Ps_1 \vdash Ps_2.$$

And it is not true that

$$(E!x) (x \in S \ \& \ Px) \vdash \overline{Ps}_1,$$

only that

$$(M^*) \vdash \overline{Ps}_1$$

This means that the argument treats (M\*) as itself part of the information mentioned by (M\*). For otherwise

$(E!x) (x \in S \ \& \ Px), (\overline{Ex}) (x \in S_2 \ \& \ Px), (M^*) \vdash Ps_2$  together with  $Ps_2$  would not be held contrary to (M\*). Hence the argument treats (M\*) as viciously self-referring. The only obvious alternative to this is that the information mentioned by (M\*) should consist of (I\*) alone. But in this case (M\*) is not strong enough to carry the inductive step.

Can we not, however, squeeze out between these two alternatives? We are given the premiss

$$(I^*) \quad (E!x) (x \in S \ \& \ Px)$$

Given also that  $Ps$  is unexpected, we may add

(M<sub>1</sub>\*) *From (I\*) it is not possible to determine which element of  $S$  has  $P$ , even with an additional premiss of the form*

$$(\overline{Ex}) (x \in S^* \ \& \ Px)$$

where  $S^*$  is a subset of  $S$  such that

$$(y) [(y \in S \ \& \ Py) \supset (z) (z \in S^* \supset z <_R y)]$$

Since, for  $S_1 = S - \{s_1\}$

$$(I^*), (\overline{Ex}) (x \in S_1 \ \& \ Px) \vdash Ps_1,$$

the proposition (M<sub>1</sub>\*) is false where  $Ps_1$ . Hence

$$(I^*) \quad (M_1^*) \vdash \overline{Ps}_1$$

It is clear that for every natural number  $n$ , we may formulate the proposition

(M<sub>n</sub>\*) *From (I\*) & (M<sub>1</sub>\*) & ... & (M<sub>n-1</sub>\*) it is not possible to determine which element of  $S$  has  $P$  even with an additional premiss of the form*

$$(\overline{Ex}) (x \in S^* \ \& \ Px)$$

where  $S^*$  is a subset of  $S$  such that

$$(y) [(y \in S \ \& \ Py) \supset (z) (z \in S^* \supset z <_R y)]$$

Now consider the principle

$$(M_n^*) \quad (M_1^*) \ \& \ [(M_n^*) \supset (M_{n+1}^*)]$$

If we adopt this principle then we will have (M<sub>n</sub>\*) for every natural number  $n$ . This will allow us to repair the gap in the fallacious argument above.

We have already proved that  $\overline{Ps}_1$ , employing (M<sub>1</sub>\*). Suppose now that we have proved each of (M<sub>1</sub>\*), (M<sub>2</sub>\*), ..., (M<sub>n</sub>\*) from (M<sub>0</sub>\*) and used these results to establish  $\overline{Ps}_1, \overline{Ps}_2, \dots, \overline{Ps}_n$  respectively. We can prove (M<sub>n+1</sub>\*) by appealing to (M<sub>0</sub>\*). And we can argue in the following way. It is obvious that for  $S_{n+1} = S - \{s_1, s_2, \dots, s_{n+1}\}$

$$(i) \quad (I^*), (\overline{Ex}) (x \in S_{n+1} \ \& \ Px), \overline{Ps}_1, \overline{Ps}_2, \dots, \overline{Ps}_n \vdash Ps_{n+1}$$

By the hypothesis of induction

$$(ii) (I^*) \quad (M_1^*) \vdash \overline{Ps}_1$$

$$(2^*) \quad (M_2^*) \vdash \overline{Ps}_2$$

.

.

(n\*)

$$(M_n^*) \vdash \overline{Ps}_n$$

From (i), (ii) we have

$$(iii) (I^*), (\overline{Ex}) (x \in S_{n+1} \& Px, (M_1^*), (M_2^*), \dots, (M_n^*) \vdash Ps_{n+1}$$

So that  $(M_{n+1}^*)$  is false for the case  $Ps_{n+1}$ . Hence

$$(n+1^*) \quad (M_{n+1}^*) \vdash \overline{Ps}_{n+1}$$

Since, for any natural number  $n$ ,

$$(M_\omega^*) \vdash (M_n^*)$$

by the principle of mathematical induction, we have now established  $(x) (x \in S \supset \overline{Px})$  upon  $(M_\omega^*)$ . This is equivalent to  $(\overline{Ex}) (x \in S \& Px)$  which is incompatible with  $(I^*)$ . The argument used is valid. The conjunction  $(I^*) \& (M_\omega^*)$  yields the contradiction

$$(Ex) (x \in S \& Px) \& (\overline{Ex}) (x \in S \& Px)$$

But there is no bite left in the paradox. For given that  $(I^*)$ , then  $(M_\omega^*)$  is unproblematically false. If  $(I^*)$ , then either  $Ps_1$  or for some natural number  $n$ , we have  $\overline{Ps}_1 \vee \overline{Ps}_2 \vee \dots \vee \overline{Ps}_n \& Ps_{n+1}$ . In the first case, for  $S_1 = S - \{s_1\}$

$$(I^*), (\overline{Ex}) (x \in S_1 \& Px) \vdash Ps_1$$

This falsifies  $(M_1^*)$  and hence  $(M_\omega^*)$ . In the second case, for  $S_{n+1} = S - \{s_1, s_2, \dots, s_{n+1}\}$

$$(I^*), (\overline{Ex}) (x \in S_{n+1} \& Px), \overline{Ps}_1, \overline{Ps}_2, \dots, \overline{Ps}_n \vdash Ps_{n+1}$$

We know already that

$$(M_i^*) \vdash \overline{Ps}_i \quad (1 \leq i \leq n)$$

Hence,

$$(I^*), (\overline{Ex}) (x \in S_{n+1} \& Px), (M_1^*), (M_2^*), \dots, (M_n^*) \vdash Ps_{n+1}$$

This falsifies  $(M_{n+1}^*)$ . But having established above that for any natural number  $n$

$$(M_\omega^*) \vdash (M_n^*)$$

we may derive as a special instance

$$(M_\omega^*) \vdash (M_{n+1}^*)$$

This means that  $(M_\omega^*)$  is falsified in the second case.

It is true that if we read " $P$  is unexpected" as  $(M_\omega^*)$  then for no element  $s \in S$  can we have " $Ps$ , and  $P$  is unexpected." But this is simply because  $(Ex) (x \in S \& Px)$  implies the negation of  $(M_\omega^*)$ . There is here no reason to suppose that there can be an element of  $S$  which has  $P$  and has  $P$  unexpectedly in the required sense. On the contrary, we have a proof that there can be no such element.

At this stage we might enter a generalization of Nerlich's second objection.  $P$  can be unexpected after all where  $(I_\omega^*)$  is true. For suppose that, say,  $Ps_{20}$ . Then it is not possible to determine from  $(I^*) \& (M_\omega^*)$  which element of  $S$  has  $P$ . Again this is either to treat  $(M_\omega^*)$  as viciously self-referring, or it is to say that  $Ps_{20}$  is unexpected in that the following proposition is true.

$(M_{\omega+1}^*)$  From  $(I^*) \& (M_\omega^*)$  it is not possible to determine which element of  $S$  has  $P$  even with an additional premiss of the form

$$(\overline{Ex}) (x \in S^* \& Px)$$

where  $S^*$  is a subset of  $S$  such that

$$(y) [(y \in S \& Py) \supset (z) (z \in S^* \supset z <_R y)]$$

The propositions  $(M_\omega^*)$  and  $(M_{\omega+1}^*)$  are distinct. The case  $Ps_{22}$  (like the case  $Ps_1$ , or any other case) provides a model for  $(M_{\omega+1}^*)$ . But no case at all where an element of  $S$  has  $P$  provides a model for  $(M_\omega^*)$ .

New College, Oxford

## VII. DELIBERATION AND FOREKNOWLEDGE

RICHARD TAYLOR\*

DELIBERATION is often confused, particularly in discussions of free will, with speculation and reasoning concerning one's future behavior. It has even been suggested that unless one could infer from certain things—e.g., from his intentions or whatnot—what he was going to do and unless, accordingly, determinism were true, then one would have no way of knowing what he was going to do—as if statements of the form “I am going to do A” were all just predictions.<sup>1</sup>

I want to make clear the great difference between deliberation, on the one hand, and speculation and inference, on the other, by eliciting some of the things that are involved in the former but not in the latter. Some of these appear to have important consequences for the “free will” controversy. It is not, however, my purpose to defend any theory of free will.

I shall go about this by listing some of the things that appear to be involved in deliberation and which distinguish it from everything else, illustrating these with examples as I go along. I shall begin with the more obvious things and conclude with the more controversial.

### I

One cannot deliberate about anything except his own possible future actions, though one can speculate or make inferences about almost anything he likes.

With respect to acts of other people, for instance, one can speculate about them, try to predict them, or to infer what they are going to be; but, one cannot deliberate about them. A statement such as “I am deliberating whether Jones will do E” cannot be true, unless it means “I am deliberating whether I shall have Jones do E,” in which case it expresses deliberation about one's own possible future act. The reason for this is that one can deliberate only about what he believes to be within his own power. Thus, “I am deliberating whether

Smith shall be reprieved” entails “I believe it to be within my power alone to reprieve Smith.” If I believe this to be within the power of another—the governor, for example—then I can speculate about what he will do, or I can deliberate about what I would do if I were the governor; but I cannot deliberate about what will be done.

Even in case of my own acts, moreover, I cannot deliberate about what I have already done or am already doing. I can deliberate only about my possible *future* acts. With respect to things I have already done, I can regret them, take satisfaction in them, and so on. If I have forgotten what those acts were, I can try to find out, infer, or guess; but I cannot deliberate about them. Though I may not know, for example, whether I took my vitamin pill yesterday, I can no longer deliberate about whether or not to take it *then*. There is simply nothing there to decide and, besides, past and present things, even if they are my own acts, are not within my power to do or to forego, and I can deliberate only about things which are. Similarly, if I am sitting, I cannot deliberate about whether to be sitting. I can only deliberate about whether to remain sitting; and this has to do with the future.

Now it would not, to be sure, be outrageously incongruous for one to say that he is deliberating or (synonymously) trying to decide whether he ought to have done something which he has in fact done, which might seem to render doubtful the claim that deliberation is concerned only with the future. Deliberation in this sense, however, is both logically and psychologically different from what I am here concerned with. It is essentially no different from what a meteorologist would be doing if, studying his data and charts, he truly said that he was trying to *decide* what tomorrow's weather is going to be, or what a moralist would be doing if he truly said he was deliberating or trying to decide whether, say, Socrates should have taken the hemlock. In such cases one is, obviously, doing nothing more than trying to resolve a question or

\* This paper (revised) was initially presented at the meetings of the Eastern Division of the American Philosophical Association, New York City, December 28, 1962.

<sup>1</sup> See e.g., J. M. E. McTaggart, *Some Dogmas of Religion* (London, Edward Arnold and Co., 1930), pp. 182-184, and R. E. Hobart, “Free will as Involving Determinism and Inconceivable Without It,” *Mind*, vol. 43 (1924), pp. 1-27.

doubt of one kind or another. The meteorologist, unless he happens also to be a rainmaker, is not trying to decide whether to have it rain tomorrow, since this is not within his power, nor is the moralist trying to decide whether to have Socrates drink the hemlock. Similarly, in deliberating or trying to decide whether I ought to have done what I in fact did, I am not trying to decide whether to do it or not, it being no longer within my power to alter that fact. I am trying only to resolve a doubt, which in this case happens to be a moral one, and what I am doing is essentially no different from what the moralist, pondering Socrates behavior, is doing. Thoughts and reflections which are aimed merely at the resolution of doubt, however, are essentially speculative rather than deliberative. When, unlike such cases, I am deliberating whether I ought to do something, which it is within my power to do or to forego, I am *not* merely trying to resolve a doubt or settle my opinion about something. Unlike the meteorologist who reflects about the weather, or the moralist who reflects upon the moral implications of Socrates' behavior, I *am* trying to decide whether to do something, or whether to leave it undone. I am trying, not merely to settle upon certain opinions, moral or otherwise, concerning what I do, but to decide just what it is that I shall do. Whatever may be the permissiveness of "ordinary usage," it is *this* which I prefer to call deliberation, in the strict sense, just to distinguish it from all those thoughts and reflections which are essentially intellectual and speculative.

Again, one cannot deliberate about such things as the future behavior of some heavenly body, even though this may be unknown to him, though he may make inferences or speculations concerning such things. One reason for this is that such things occur by necessity, as Aristotle pointed out, and are not within anyone's power to control. But that is not the only reason. One could no more deliberate about, say, the outcome of the spin of a roulette wheel, even if he assured this to be casually undetermined. He could only guess, make bets on it, and so on—unless, of course, he thought he could influence this outcome. But then he would be deliberating on his own future activity—namely, whether or not to try influencing this outcome.

Finally, I have said that deliberation is concerned with one's *possible* future actions, and this is a qualification that is dictated by the pre-conditions of deliberation as well as by logic. As we shall see shortly, an action which is believed to be inevitable

can be no subject of deliberation nor, by the same token, can one which is believed to be impossible. Beyond that, however, if one is deliberating concerning certain *alternative* actions, then not all of them can be, simply, his future actions. Each can be no more than a *possible* action. If, for example, I am deliberating whether to leave the room or to stay, then not both of these can be my future actions, for on the supposition that either of them is my future action, it logically follows that the other is not.

## II

One cannot deliberate about his own future act, in case he believes the act in question is already inevitable.

This is, again, a consequence of the fact that one can deliberate only about what he believes to be within his power to do and to forego, and the very point in calling anything inevitable is to deny that this condition exists. Thus, one cannot deliberate about whether to (eventually) die; he can only deliberate on how to make the best of it, with insurance and so on. The husband of a pregnant woman cannot deliberate on whether to become a father, unless this is a question of whether to terminate the pregnancy. A passenger in an airplane cannot deliberate about whether or not to return to earth; he well knows that he will, in one way or another. He cannot even deliberate about when or where to come down, unless he is the pilot—i.e., unless this is up to him, or within his power. In case such things are thought not to be "acts," we can add that a soldier cannot deliberate about whether or not to arm himself, in case he knows that there is a regulation requiring him to do so, and that the regulation will be enforced. What to do is, in this case, not up to him.

Now of course one can deliberate whether to do this or that *if* a certain condition is fulfilled, not knowing whether that condition will be fulfilled but believing that it has already been rendered inevitable that it will be, or that it will not. One might, for example, deliberate whether to study in France or in Italy in case he gets a certain award, knowing that the awards have already been finally decided but not yet announced. In that case he can only guess, speculate, or even try by secret intelligence to find out whether he has won an award. But without doing any of this he can still deliberate about whether to go to France or to Italy, in case he does get it. In that case, however,

he must believe that neither of these two alternatives is likewise already rendered inevitable, in case he has won the award. He cannot, for example, believe that the award, in case he has won it, will turn out to be one permitting him to study only in Italy, or only in France, and still deliberate about where to study on the award. At most he can then only deliberate about whether or not to accept the award, in case he turns out to have won it.

### III

One cannot deliberate about what he is going to do, even though this may be something that is up to him, at the same time knowing what he is going to do.<sup>2</sup>

This is one thing that deliberation has in common with speculation, inference, and guesswork; namely, that all presuppose ignorance, in the absence of which they can only be shammed. Inference about things future, however, has for its purpose the *discovery* of what is *going* to happen, whereas deliberation, which is necessarily about things future, has for its purpose a *decision* or "making up one's mind" about what to *make* happen, and in this respect the two are utterly different.

There seem, in fact, to be only these two ways in which one could know what he is going to do; namely, by *inferring* what he is going to do, or by *deciding* what he is going to do. In neither case can one deliberate about what he is going to do.

Thus, if a governor said "I am, as a result of my forthcoming deliberations, going to reprieve Smith," he would indicate that his mind was already made up, and hence, that he was not going to deliberate about it—unless, of course, with a view to possibly changing his mind. But in that case he could not know that his statement was true. He could, of course, pretend to deliberate about it, discuss the matter with his assistants, perhaps publicly review the pros and cons once again, but if he did so he would be shamming deliberation. His purpose would not be to arrive at a decision, this having been already arrived at, but something else—perhaps that of conveying a desirable public image of himself.

Similarly, if anyone said "I see, by reliable signs and portents, that I am about to do E, so I shall deliberate about it," he could not possibly be

expressing himself accurately. If he does already know what he is going to do, there is nothing there for him to decide, and hence nothing to deliberate about.

For example, it might be possible for a group of observers to infer reliably from certain signs that a certain man is about to be married. They see the flowers, witnesses assembled, preacher waiting, music playing, groom suitably attired, and so on. From the same evidence, which is apparent to the groom himself, he too can gather that he is about to be married, though for him, unless he doesn't realize what he has gotten himself into, such signs and portents are superfluous. If, however, he regards these signs as reliable evidence of what he is about to do, he cannot deliberate about what to do—he is past deliberation, and the die is cast. If, on the other hand, he still does deliberate about whether to get married—if he has last minute misgivings and second thoughts—then he obviously does not regard the signs as reliable evidence of what he is going to do. He is, in fact, contemplating confuting the very thing those signs point to, by walking right out of the church.

Of course deliberation is seldom if ever so pure as this. More commonly one finds himself partly trying to decide what to do, partly trying to predict what he is going to do, partly deliberating about what to do if the predictions turn out right and, perhaps in addition, partly deliberating about whether to hold to a decision that has been at least tentatively made, and so on. Mixed with our governor's deliberations, for instance, might be all sorts of attempts at predicting what his opponents will do, what he will be forced to do in response, and what, in the light of these, he ought to do about this reprieve, and so on. Still, deliberation about what *to* do is essentially different, both logically and psychologically, from prediction about what one is *going* to do, or what other people or things are going to do. One can deliberate, but not predict, about what to *make* happen, and one can predict, but not really deliberate, about what is *going* to happen. The fact that both can occur together and have significant connections with each other, and are for this and other reasons often confused in the minds of philosophers and others, does not obliterate the essential differences at all.

<sup>2</sup> This point is derived from Carl Ginet's paper, "Can the Will be Caused?" *Philosophical Review*, vol. 71 (1962), pp. 49–55. See also Stuart Hampshire and H. L. A. Hart, "Decision, Intention and Certainty," *Mind*, vol. 67 (1958), pp. 1–12, referred to by Ginet.

## IV

If one's act is caused, in the usual sense—i.e., is the inevitable consequence of certain conditions existing antecedently—then he can, simply by his awareness of those causes, know by inference what his act is going to be.<sup>3</sup>

Under such circumstances—i.e., the awareness of such causes and knowledge of their consequences—one cannot, of course, deliberate whether to do the thing in question, for he already knows that he will. Examples are supplied by compulsions, addictions, solemn agreements, and the like. Or consider some such act as sneezing, which is ordinarily performed involuntarily but which can be done deliberately. If one feels a sneeze coming on, in the sense that he is forewarned of this impending convulsion by a certain familiar nasal tickle, then he cannot deliberate whether to sneeze or not; he can only prepare for it. The only exception would be in case he thought he might be able to repress the sneeze; but in that case he would not, obviously, consider the felt irritation to be causally sufficient to make him sneeze. One might, on the other hand, have some occasion to deliberate whether to sneeze, if he were considering ways of attracting someone's attention, for example, or perhaps of feigning illness in order to avoid some irksome chore. His deliberation would have to cease, however, the moment he became aware of any condition sufficient either for his sneezing, or for his not sneezing, for he would then know what he was going to do.

From this it of course follows that one's deliberate acts cannot be caused, in the usual sense, or, if they are, then he cannot know that those causes exist at the time he deliberates. Like speculation about what is going to happen, then, deliberation about what is going to happen, or, more precisely, about what one is going to make happen, rests upon ignorance.

But now the question arises whether deliberation rests upon anything more; that is, whether it presupposes only an *ignorance* of the causes of one's deliberate act, or the actual *absence* of such causes. We shall return to this important question shortly, but here we can note that it is quite possible for one to deliberate about whether to do a certain thing even in the presence of conditions causally sufficient for his doing what he contemplates doing, provided, of course, that he is ignorant of the existence of such conditions. One might, for instance, be deliberating

whether to sneeze, thinking that this might be an effective way of feigning illness, not knowing that a sneezing powder has been liberated into the room, the inevitable effect of which will soon be to cause everyone in the room to begin sneezing. Or one might be deliberating whether to leave a certain house, wholly unaware that the house is on fire and he will shortly be forced to leave. One can hardly help noting, however, that in such cases one's deliberation is otiose and pointless, since what one then does is not the *result* of his deliberation at all. There was really nothing for him to decide; he only thought there was.

## V

If one does not know what he is going to do, but knows that conditions already exist sufficient for his doing whatever he is going to do, then he cannot deliberate about what to do, even though he may not know what those conditions are.

One can, in such a case, only guess or speculate about what he will do, or try to find out what it is that he will be forced to do. This is a consequence of the fact that one can deliberate whether to do a certain act only if he believes it is up to him whether to do it or not, or, that it is within his power equally to do it, and to forego it.

For example, consider a soldier who knows that daily orders regarding the bearing of arms are enforced, and that he has no choice but to obey them. Suppose he does not know whether or not he shall be required to arm himself today, though he knows that the order has been posted. He cannot deliberate about whether to arm himself today. He can only check to see what order has been posted and, until then, perhaps try to guess. Of course he might deliberate whether to comply with his order; but if he did he would not be assuming that such orders are really *enforced*. He would be assuming only that there are strong, but perhaps insufficient, inducements for obedience.

Or consider a man—we'll call him Adam—who has spent the evening at the distant home of a friend and is then invited by his host to spend the night. This might call for careful consideration of the pros and cons on Adam's part, for weighing in his mind the pleasures of staying over as against considerations of his responsibilities at home, and so on. Suppose further, however, that another guest—we'll call him Brown—knows that there exist conditions which render it causally impossible

<sup>3</sup> Ginet, *op. cit.*, p. 50.

for Adam to go home. He knows, for instance, that the last train has left, and that there is no other way for Adam to get home. Now clearly, Adam can still deliberate about whether to remain or not, in ignorance of what Brown knows. But now suppose Brown announces that he knows what Adam is going to do, without giving any hint as to what this is, and that he knows it on the basis of certain unnamed conditions which are causally sufficient for Adam's doing what Brown knows he will do. If Adam *believes* this, he cannot any longer deliberate about what to do, even though he does not know what he is going to do and is not himself aware of any conditions sufficient for his doing either the one thing or the other. All he can do is speculate, guess, and wait to see what he will have to do, meanwhile exhorting Brown to tell him. He can no longer deliberate about the matter because, if he believes Brown, then he believes it is not up to him what he does; the matter has already been "decided," one way or the other, and there is no decision for Adam to make.

It is no good here, incidentally, to introduce such vague and familiar slogans as "Deliberation might, after all, be a natural process," or "Deliberation is only the way some, perhaps psychological, causes work themselves out," and so on. If such remarks are unpacked, and "natural processes" are found to be nothing but causal chains, and "causes" are understood to be causes of the usual kind—namely, antecedent conditions, psychological or other, which are sufficient for, and thus render inevitable, whatever it is that they cause—then far from being rejoinders to what has been said they only illustrate something that is painfully well known; namely, that philosophers, no less than the vulgar, are perfectly capable of holding speculative opinions that are inconsistent with some of their own beliefs of common sense.<sup>4</sup>

Now I believe the principle involved here can be generalized, such that if a man believes that there are, or ever will be, conditions, not themselves within his control, sufficient for his doing whatever it is that he is going to do, then he cannot deliberate about what to do, even though he may not have the slightest idea what this is, or the slightest idea

what those conditions are, or will be, or what they will be sufficient for.

Consider a man at a cocktail party, for instance, who knows, in a cognitive sense of "knows" which entails that what he knows is true, that he will accept any standard cocktail that is offered provided it is made with gin, but that he will drink nothing alcoholic otherwise, having a nausea for any other type of spiritous beverage. Now this man cannot deliberate about whether to drink gin, for he already knows that he will, *if* it is offered. There is, then, nothing there for him to decide. Nor can he deliberate about whether to drink at all, for he already knows that he will not, *unless* gin is offered, so there is nothing there to be decided. All he can do is try to speculate, or guess, whether gin will be served, this being, we are supposing, something that is not up to him. And it should be noted that under the conditions assumed it is impossible for him to deliberate, even though he may not know what he is going to do, and may even doubt that conditions already exist which are sufficient for his doing whatever he is going to do.

This example is imperfect, however, for one can justly wonder how anyone could have such knowledge. One can "know" what he is going to do under certain and as yet undecided alternative circumstances, in the sense of having firmly made up his mind—and still, for instance, fall dead before having a chance to do it, showing that his "knowledge" was not of the kind that entails that what was thus "known" was true. This observation does not really affect the argument, but since the doubt raised about the illustration can easily transfer itself to the argument we should perhaps supply a better example. Consider, then, a man who is watching the spinning of a roulette wheel, and who knows (and has not merely resolved) that he will take the purse in case it stops on an even number, but that he will have to surrender his own stake in case it stops on an odd number. Now he cannot deliberate about whether to take the purse or surrender his own, even though this has not been at all determined. And, it should be noted, this is still true, even if he believes the behavior of the wheel to be causally undetermined with respect to where it stops, and hence believes that his own act

<sup>4</sup> A well-known philosopher is alleged to have announced to an audience that he was a solipsist, and that he could not understand why they were not all solipsists too, and I once heard a philosopher claim that he knew nothing at all, not even that he was enunciating that opinion to me. These are extreme examples of the kind of muddle some philosophers have appeared to me to be involved in when they have said that they are determinists who deliberate—as if this were some sort of challenge or rejoinder to something. Merely pointing out that certain views are held, even by philosophers, or even by oneself, does not prove that the views are consistent and is sometimes a *prima facie* reason for suspecting they are not.

will have been causally undetermined as well, such that there are not yet any conditions sufficient either for his doing the one thing, or for his doing the other. The reason for this is obvious; namely, that having got this far into the game it is no longer up to him what he does. It is entirely up to the roulette wheel, and there is nothing for him to decide. All he can do is guess, and hope.

Now we can, I believe, extend this principle still farther, and say that if a man knows that there will at any future time be some condition sufficient for his having done a certain act in the meantime, then he cannot deliberate whether to do that act, even though he does not know what his act will be or what that condition will be.

This is, of course, plainly false on one natural interpretation, for one sometimes knows that, whichever of two alternative things he does, there will then be traces from which it can be certainly inferred what he has done, and he can, nevertheless, deliberate about which thing to do.

Nevertheless, if one were to learn that there was going to be a certain condition, as yet unspecified, for his doing a certain act in the meantime, then he could not deliberate about that act, even though he did not yet know which act it is. He knows, or can infer, that he can act only in whatever way is necessary for the occurrence of that condition, whatever it is.

For example, suppose a man knows that if he is found in a certain place at some given future time, this will be sufficient for his having gone there in the meantime, whereas if he is found in another place at that time, this will be sufficient for his having gone to that other place. He does not know where he will be found, and hence, where he will have gone in the meantime. There is nothing so far, then, that prevents his deliberating, with a view to making up his mind, where to go, and where, accordingly, to be found. But now suppose he learns that some other person does somehow already know where he will be found—not that this other person has a fair idea or can make a more or less educated guess, but that, somehow or other, he actually knows. Now I believe the first man cannot, if he knows that another man is in possession of such knowledge, any longer deliberate about where to go; he can only wait and see where he is going. It is not within the power of any man to

render false what another man knows to be true.<sup>5</sup> To the extent that he *can* deliberate about where to go, to that extent he must consider it doubtful that anyone could already know where he will subsequently be found. One deliberates, not about what *will* happen, but about what to *make* happen; and if it is up to him what he shall make happen, then it is also up to him what shall eventually be true about what has happened.

## VI

No one can know what another is going to do as a result of forthcoming deliberation, nor could God have such foreknowledge.

One can, of course, know what another is going to do as a result of deliberation that is already concluded, for that person can then simply announce what he is going to do. But one can make no such announcement while still deliberating, for he could not himself know that it was true. There is no way that he could possibly know, before he has decided, nor is there any way that anyone else could know.

If someone knew what another was going to do as a result of forthcoming deliberation, then he would know on the basis of some kind of evidence; that is, on the basis of his knowledge of certain conditions that were sufficient for the agent's doing the thing in question, and from which it could be inferred that he would do that. But if there were such conditions, then they could also be known by, or made known to, the agent himself, such that he too could infer what he was going to do. This, however, is impossible, so long as the agent has not yet himself decided what to do. Indeed, the agent cannot even believe that any such conditions, known or unknown, exist, and at the same time believe that it is within his power both to do, and to forego doing, the thing in question. This, as we have seen, appears to be a necessary condition of deliberation.

The foregoing is not to be confused with a familiar type of fallacy, whereby one truly asserts what cannot happen in case something else happens, and then, ignoring this qualification, draws some categorical conclusion about what cannot happen. The point is rather, that no one can know by inference that a certain event is going to happen,

<sup>5</sup> There is room for endless misunderstanding in this statement, but one must try to resist the temptation to say that while it is within a man's *power* to render false what another knows to be true—meaning by that only that what is known to be true could be false—no man ever *does* render false what another knows to be true. That familiar modal fallacy, taught to all philosophy students early in the game, is not involved here.

except on the basis of his knowledge of certain conditions sufficient to produce that event. If no such conditions exist, then it obviously cannot be known by inference that the event in question is going to happen, and if it is so known, then there must be such conditions. If the event in question is the act of some agent, however, then that agent cannot deliberate about whether to do it, believing that any such conditions already exist, even though he may not know what they are; for the fact that *any* such conditions already exist would entail that it is no longer up to him what he is going to do. And moreover, if another person knows by inference what his act will be, then he cannot know that this act will be the result of deliberation still forthcoming. He will, on the contrary, know that it will be the result of conditions, known by him, sufficient to produce it.

Suppose, for example, that I feel confident that a certain man, now deliberating whether to go to Boston or to New York, is going to decide to go to New York. Now if I am really confident of this, and my confidence rests upon something more than a mere feeling or hunch, then I cannot believe that he is really deliberating with a view to deciding where to go. I must instead believe that the matter is already fairly settled in his own mind, and that he is, at best, only reviewing the pros and cons of what he has already fairly decided. If, on the other hand, I believe that he is really deliberating about where to go, in the sense which presupposes that it is up to him where he goes, then I cannot feel confident that he will go to either place rather than the other. What he finally decides is something that is up to him, if his decision is really the result of his deliberation.

Now I might, to be sure, know a person and his habits well enough to know that, whenever he is confronted with a certain choice—say, that of going to New York or to Boston—then he invariably decides the same way—say, by going to New York. And it is possible to suppose that, before deciding, he always or often deliberates about the matter. In that case I could predict with confidence what he was going to do, and this would be consistent with his always deliberating first. But then I would know what he was going to do, *not* as a result of his deliberation, but as a result of something else—of habit, for example. If, as a result of sheer habit, or as a result of some other condition that is always present when such a decision is made, the man invariably decides in the

same way, then his decision is not the result of his deliberation, and not something that is really up to him. It is the result, or causal consequence, of something else; of habit, for example, or of whatever other condition we are supposing determines the matter.<sup>6</sup>

If, moreover, I know that another person is deliberating about a certain choice that is before him, and know what his decision is going to be, on the basis of some consideration that is known to me and which must sooner or later also come to his attention and certainly decide the matter, then I know what he is going to do, not as a result of his deliberation, but as a result of this further consideration, which will terminate his deliberation. If, for instance, I know that someone is deliberating whether to remain in the room or leave, and I know, further, that the room is on fire, and that he will shortly notice this himself, and leave as a consequence of this, then I know what he is going to do. But I know this only because I know that what he is going to do will *not* be the result of his deliberation, but of his knowledge of the circumstances. What he does is not up to him at all, assuming the fire to be of such a nature as leaves him no real alternative.

From the foregoing it becomes apparent that the inability of anyone to know what someone is going to do, as a result of deliberation, does not result from any limitations of human sagacity, but from the very nature of a deliberately chosen act. God, accordingly, can have no more such foreknowledge than any man; and it is no rejoinder to this simply to *define* God as an omniscient being.

If God had foreknowledge of the deliberate act of some man, then that knowledge could be shared with that man himself. At least, there is no reason why it could not. But that is impossible, for no man can continue to deliberate about whether to do something, if he already knows or can know what he is going to do. There is an absurdity in the conception of a man learning (as contrasted with merely hearing) from God that he is going to deliberate about whether to do or forego doing a certain act, and then, as a result of his deliberation, that he is going to forego it. Nor is this just a consequence of the fact that, if a man is deliberating about what to do, then it would be a contradiction to say that he already knows what he is going to do. The fact that a man is deliberating is no *obstacle* to his knowing, or learning, anything whatever, any more than a man's being a bachelor is an obstacle

<sup>6</sup> See John Canfield, "Knowing about Future Decisions," *Analysis*, vol. 22 (1962), pp. 127-129.

to his marrying someone. The fact that a man knows or can find out by inquiry what he is going to do, on the other hand, is an obstacle to his deliberating about it and then doing it as a result of such deliberation, just as the fact that a man has a wife is an obstacle to his having still another.<sup>7</sup>

Even if such supposed divine knowledge could not, for some reason, be shared with men, it still could not exist, even for God. For to the extent that a man can deliberate whether to do one act or another, to that extent he believes that each act is equally within his power. Hence, on the supposition that God believes the man will do the first of these acts and not the second, or that he will do the second and not the first—whichever of these suppositions one chooses—the man must also believe it to be within his power to confute God's belief, which is absurd. This is not, it should be

noted, to say that foreknowledge, whether human or divine, is ever any cause by itself, or that it exerts any compulsion on anything whatever.

## VII

There can be no truth or falsity in any assertion about what any man's future deliberate act will be.

This, combined with the supposition of God's omniscience, is a consequence of what has already been said. For if God is omniscient, then he knows everything that can be known, which is for God exactly coextensive with everything that is true. Hence, if as between the assertion and the denial that a man will, as a result of deliberation still forthcoming, do a certain act, God cannot know which is true, it follows that neither is true, and accordingly, that neither is false.

<sup>7</sup> Any reader to whom the point of this analogy is unclear may ignore it.

# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson  
Kurt Baier  
Richard B. Brandt  
Roderick M. Chisholm  
L. Jonathan Cohen  
James Collins  
Peter Thomas Geach  
Adolf Grünbaum

Norwood Russell Hanson  
Carl G. Hempel  
Jaakko Hintikka  
Jean Hyppolite  
Raymond Klibansky  
Helmut Kuhn  
Benson Mates  
John A. Passmore  
Günther Patzig

Oliver L. Reiser  
George A. Schrader  
Wilfrid Sellars  
J. J. C. Smart  
Wolfgang Stegmüller  
John Wild  
G. H. von Wright  
John W. Yolton

---

VOLUME 1/NUMBER 3

JULY 1964

## CONTENTS

- |  |     |   |     |
|--|-----|---|-----|
| I. HANS AARSLEFF: <i>Leibniz on Locke on Language</i> . . . . .    | 165 | IV. ISAAC LEVI AND SIDNEY MORGEN-<br>BESSER: <i>Belief and Disposition</i> . . . . .      | 221 |
| II. CHARLES H. KAHN: <i>A New Look at Heraclitus</i> . . . . .     | 189 | V. K. W. RANKIN: <i>Referential Identifiers</i> . . . . .                                 | 233 |
| III. ROBERT C. COBURN: <i>Recent Work in Metaphysics</i> . . . . . | 204 | VI. GEORGE SCHLESINGER: <i>The Problem of Evil and the Problem of Suffering</i> . . . . . | 244 |
-

# AMERICAN PHILOSOPHICAL QUARTERLY

## POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

## SUBSCRIPTIONS

The individual subscription price is five dollars *per annum* for individuals (four dollars for students and for subscribers not resident in the U.S.A. or Canada) and ten dollars *per annum* for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. While available, back issues will be sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscriptions and back orders may be addressed directly to the publisher (University of Pittsburgh Press, University of Pittsburgh, Pittsburgh, Pennsylvania 15213).

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with a wide margin, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively—also being typed with wide margins and spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## REPRINTS

Authors who are subscribers will receive 50 reprints gratis. Additional reprints can be purchased through arrangements made when checking proof.

\* \* \*

The *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.

Application to mail at second-class postage rates is pending at Pittsburgh, Pennsylvania.



# I. LEIBNIZ ON LOCKE ON LANGUAGE

HANS AARSLEFF\*

THE subject of this paper is Locke's critique of language and Leibniz' reaction to that critique. The basic texts are Locke's *Essay* and Leibniz' discussion of its contents in the long dialogue which he gave the title *Nouveaux Essais sur l'Entendement*.<sup>1</sup> In spite of agreement on some points, their positions are fundamentally opposed. My purpose is to define the positions of each and to attempt to explain why Locke's analysis of "Words or Language in General"—as he phrased it in the title of III, i—proved unacceptable to Leibniz. My thesis is that their basic opposition turns on the question whether language is conventional or natural, and I shall try to show that Locke had deliberately taken his stand against the doctrines which Leibniz later advanced in his critique. Thus their disagreement can be defined within the

history of the lively seventeenth-century debate about words and things, language and mind, and ultimately language and knowledge. Or, put differently: If the object of man's study is Nature, as to the Royal Society it certainly was, does language somehow find its place within Nature or does it exist only for man's convenience and entirely apart from Nature? If the former is the case, then one may expect that words can be made to yield some—and perhaps much—knowledge of reality; if the latter, none; and in either case, language may perhaps reveal something about the mind. My approach is historical, not philosophical; and I shall leave aside the aspects of Leibniz that have recently attracted the most attention, i.e., the logic, the calculus, the combinatorial characteristics and other related matters that

\* I wish to thank the Princeton University Research Committee for summer grants in 1962 and 1963. I have here used a small part of the research they made possible. I especially wish to thank Professor William G. Moulton for opening the forum of the Linguistics Discussion Group to me; this paper was first read to that group in May of 1962, in a different and much shorter form.

<sup>1</sup> Most frequently used texts with abbreviations:

**Essay:** John Locke, *An Essay Concerning Humane Understanding*, 5th ed. (London, 1706). This was the last edition to receive Locke's own revisions. This text is now available in *Everyman's Library*, nos. 332 & 984 (London, 1961), edited with an Introduction by John W. Yolton. This edition now supersedes A. C. Fraser's edition. Fraser's edition is not reliable: The text is inaccurate, the marginal section summaries do not follow Locke's own, and it often fails to give the textual information it purports to give.

**Nouveaux Essais:** Gottfried Wilhelm Leibniz, *Philosophische Schriften*, vol. VI (Berlin, Akademie-Verlag, 1962; *Sämtliche Schriften und Briefe*, hrsg. von der Deutschen Akademie der Wissenschaften zu Berlin; ser. 6). This is the latest volume in the Akademie Ausgabe of Leibniz, this volume edited by the Leibniz-Forschungsstelle der Universität Münster. This volume contains the *Nouveaux Essais* and related pieces, with an Introduction and apparatus. (See also Gerhardt below.) All quotations from the texts of the *Essay* and the *Nouveaux Essais* are from these two editions, the 5th ed. (London, 1706) and the Akademie Ausgabe. However, I do not follow the latter's practice, stated "Introduction," p. xxvii: "Les extraits de la traduction de Coste que Leibniz met dans la bouche de Philalèthe sont imprimés en italique"; for these passages do not in all cases follow *verbatim* the Coste translation. All references are given in the text by book, chapter, and section, a procedure that will facilitate easy cross-reference to other editions in case these texts are not available.

**Locke, Works:** *The Works of John Locke*, 9 vols. (London, 1794). The text of *Some Familiar Letters* is in vol. VIII, pp. 285-472 (correspondence with Molyneux) and in vol. IX, pp. 1-145 (correspondence with Limborch).

**Aaron & Gibb:** R. I. Aaron and Jocelyn Gibb, *An Early Draft of Locke's Essay*, together with Excerpts from his Journals (Oxford, 1936). Unless otherwise noted, quotation from the Journals refer to this edition.

**King, Life:** Peter King, *The Life of John Locke, with Extracts from his Correspondence, Journals and Commonplace Books* (London, Bohn's Standard Library, 1858).

**Lough, "Locke's Reading":** John Lough, "Locke's Reading during his Stay in France (1675-79)" in *The Library*, Fifth Series, vol. 8, no. 4 (December, 1953), pp. 229-258.

**Coste:** *Essai Philosophique concernant L'Entendement Humain* par M. Locke. Traduit de l'Anglois par M. [Pierre] Coste (Amsterdam, 1742).

**Gerhardt:** *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz*, hrsg. von C. I. Gerhardt, 7 vols. (Berlin, 1875-1890). Vol. V contains the *Nouveaux Essais*, but both the Introduction and the text have been superseded by the Akademie Ausgabe listed above.

**Boyle, Works:** *The Works of the Honourable Robert Boyle* [ed. Thomas Birch], 6 vols. (London, 1772).

have been treated by Russell, Couturat, and their followers.

The contrasted views of Locke and Leibniz form the final stage in the history of a subject that can most briefly be named Language and Natural Philosophy in the Seventeenth Century, with which I hope to deal in the future. It is a large subject, but here I shall include only as much of the background as is required to identify the origins of the opposition. This procedure will, I hope, also help to clarify and fix some fundamental aspects of Locke's doctrines and method in the *Essay*, which have been either poorly understood or have received interpretations that are too diverse for comfort and conviction. The *Essay* was literally epoch-making, and such works never fail to efface their own past; in fact, one can almost say that the *Essay* has no other history than that which was its own future, as if Locke merely wrote to give Berkeley and Hume something to write about. Unlike Locke, Leibniz often named his sources and gave citations, which may help to identify the *Essay's* elusive background; and Leibniz' vast correspondence offers a further wealth of information.

For the sake of preliminary orientation, it may be useful to make a few remarks about language and philosophy, natural and otherwise, in the seventeenth century. It is familiar knowledge that many put upon language or words the burden of philosophical imperfection. Hence the success of philosophy must depend on overcoming the "cheat of words" as it was often phrased.<sup>2</sup> Two solutions offered themselves. On the one hand we find some men offering new signs modeled on the principle behind Arabic numerals, algebraic notation, and even hieroglyphics or Chinese

characters, all of which are often mentioned in this context—Leibniz among others hoped that such signs might eventually become the only vehicle of philosophical knowledge. On the other hand we find an increasing concern with plain language and words. Here the solutions were more in doubt, but no less eagerly sought to overcome the stultifying consequences of the language of the schools—of scholasticism, its terms, and its nominalism—to take only the best-known instance. Here Locke's answer was certainly the most radical and has also proved the most lasting. Both were designed to insure that the reading of the book of nature be raised above confusion, uncertainty, and controversy for the sake of peace and order and the improvement of man's estate in this world. Religious devotion demanded no less. "Empty words" must be subdued to serve "real meaning." The other book of the Creator, Scriptures, required a similar answer to remove its reading beyond controversy and such subsequent disasters as the Thirty Years' War; thus all the great natural philosophers made a special point of knowing the sacred languages: Boyle, Newton, John Pell, Robert Hooke, and Locke, to mention only a few.<sup>3</sup>

The words of two men will illustrate two separate and typical doctrines. In the *Discours*, Descartes had said: "Ceux qui ont le raisonnement le plus fort, et qui digèrent le mieux leurs pensées, afin de les rendre claires et intelligibles, peuvent toujours le mieux persuader ce qu'ils proposent, encore qu'ils ne parlent que bas-breton, et qu'ils n'eussent jamais appris de Rhétorique." This was a defiant statement, and it did not fail to elicit a commensurate reaction; for men of good sense plain words, in whatever language or dialect,

<sup>2</sup> See, e.g., Locke to Molyneux, (26 Dec. 1692): "I can easily forgive those who have not been at the pains to read the third book of my essay, if they make use of expressions that, when examined, signify nothing at all, in defence of hypotheses, that have long possessed their minds . . . I find none so fit, nor so fair judges, as those whose minds the study of mathematics has opened, and dis-entangled from the cheat of words, which has too great an influence in all the other, which go for sciences: and I think (were it not for the doubtful and fallacious use that is made of those signs) might be made much more sciences than they are." Locke, *Works*, vol. VIII, pp. 300-301.

<sup>3</sup> Cf. Thomas Birch, *The Life of the Honourable Robert Boyle* (London, 1744), pp. 98-99, quoting from "some loose sheets, intended as a part of his *Essay on the Scripture*," begun in 1652: "Those excellent sciences, the mathematics, having been the first I addicted myself to, and was fond of, and experimental philosophy with its key, chemistry, succeeding them in my esteem and applications; my propensity and value for real learning gave me so much aversion and contempt for the empty study of words, that not only I have visited divers countries, whose languages I could never vouchsafe to study, but I could never yet be induced to learn the native tongue of the kingdom I was born and for some years bred in." Cf. *ibid.*, p. 102, Boyle to John Mallet in Jan. 1652/3: "I [was] . . . very much delighted to find too, that you began to have a friendship for the Eastern tongues. For though to a person so used to the study, and replenished with the knowledge of things, as Mr. Mallet, the learning of words cannot but at first be very tedious; yet since to be a good grammarian is necessary to be a good divine; and he, that hath no skill in the original scripture himself, may be deluded by those, that translate it for him; you will find a rich compensation for the trouble of learning the holy tongue in the advantages of having learned it; and by the help of that primitive language, wherein they were written, you may gain a free and safe access to those theological mysteries, which he, that is no linguist, must either totally ignore, or take upon trust."

are enough.<sup>4</sup> To Locke the situation was somewhat different; he found that "the Extent and Certainty of our Knowledge . . . had so near a Connexion with Words, that unless their force and manner of Signification were first well observed, there could be very little said clearly and pertinently concerning Knowledge" (III, ix, 21). To Locke, unlike Descartes, the obstacle to good sense and knowledge was not merely a verbose enemy, not just some men's words, but words. The problem had become more acute since 1637 and pressed with special urgency on the experimenters in the Royal Society, who had to find good words and plain discourse to say the things they did. Leibniz would have agreed with Locke about the need for paying close attention to words, but he did not agree with Locke on the nature of language and words; hence "their force and manner of Signification" became a point at issue between the two philosophers.

# I.

It is familiar knowledge that the *Essay*, having been at least nearly 20 years in the making, made its first appearance in the world very early in 1690 and that Book III bears the title "Of Words." It is not so often remembered that both Book II and Book IV also have a good deal to say about words, especially the former. This is merely one

example of the repetitiveness of the *Essay*, which Locke was aware of, but said he was "too lazie, or too busy" to remedy.<sup>5</sup> It is also familiar knowledge that Locke at the very end of Book II made the remark that he had not originally planned to devote a separate book to words but now found that it was "impossible to speak clearly and distinctly of our Knowledge, which all consists in Propositions, without considering, first, the Nature, Use, and Signification of Language." In III, ix, 21 he further confessed, "that when I first began this Discourse of the Understanding, and a good while after, I had not the least Thought, that any Consideration of Words was at all necessary to it."<sup>6</sup> The discussion of words had assumed crucial importance in his argument. In a letter to Molyneux of 20 Jan. 1692/3, he later admitted that "some parts of that third book . . . cost me more pains to express, than all the rest of my essay."

On the basis of the statements just quoted, we are sometimes told that Locke's close attention to words was an afterthought; perhaps it was in relation to the first inception of "this Discourse of the Understanding." But in that case the inception must by "a good while" have preceded the two 1671 drafts, for both of these grant words and their meanings a prominent place in the argument.<sup>7</sup> Thus, compared with the drafts, Book III marks a change in emphasis and strategy rather than in

<sup>4</sup> In the *Principes*, in the "Lettre de l'auteur à celui qui a traduit le livre, laquelle peut ici servir de préface," Descartes made the same point in different terms: "Il faut conclure que ceux qui ont le moins appris de tout ce qui a été nommé jusques ici philosophie sont les plus capables d'apprendre la vraie."

<sup>5</sup> See "Epistle to the Reader" in the paragraph that explains the consequences of "this discontinued way of writing." On several occasions, Locke himself referred to the repetitions and would have been happy to see them removed. Writing to John Wynne at Oxford, who had proposed to do an abridgment of the *Essay*, he said: "Let me add that several of those repetitions, which for reasons then I let it go with, may be omitted, and all the parts contracted into that form and bigness you propose." King, *Life*, p. 192. See also Locke to Molyneux, 26 April 1695, regarding the Latin translation then proposed: "One thing particularly you will oblige me and the world in, and that is, in paring off some of the superfluous repetitions, which I left in for the sake of illiterate men, and the softer sex, not used to abstract notions and reasonings. But much of this reasoning will be out of doors in a Latin translation." Locke, *Works*, vol. VIII, p. 356.

<sup>6</sup> Cf. III, ix, 16.

<sup>7</sup> See "Comparative Table of the *Essay* and Drafts" in Aaron & Gibb, p. 129. Thus in par. 4 (pp. 12-13), Locke says: "And therefor in the discourse I have here made concerning humane Intellect I could not avoid saying a great deale concerning words because soe apt and usuall to be mistaken for things." See also the rest of par. 4 and the concluding part of the long par. 27 (pp. 48-49), which went nearly *verbatim* into the *Essay* as IV, viii, 10. Locke likewise shows much interest in language in the entries on "Study" that he made in his Journal during the travels in Southern France between March and May 1677. See, e.g., King, *Life*, p. 105: "Words are, in their own nature, so doubtful and obscure . . . that if, in our meditations, our thoughts busy themselves about words, and stick at the names of things, it is odds but they are misled or confounded. This, perhaps, at first sight may seem but a useless nicety, and in the practice, perhaps, it will be found more difficult than one would imagine; but yet upon trial I dare say one's experience will tell him it was worth while to endeavour it." See also two journal entries on "Species" given in Aaron & Gibb, pp. 83-84 (19 Sept. 1676, written at Montpellier) and pp. 98-99 (19 Nov. 1677, written in Paris); the latter says: "The species of things are distinguished and made by chance in order to naming and names imposed on those things which either the conveniencies of life or common observation brings into discourse." This is the subject that Locke deals with in III, v and especially in III, vi, 7ff, and the verbal similarities are striking; cf., e.g., III, vi, 36: "'Tis Men, who, taking occasion from the Qualities they find united in them, and wherein, they observe often several individuals to agree, range them into Sorts, in order to their naming, for the convenience of comprehensive signs," and III, vi, 37: "So that we may truly say, such a manner of sorting of Things, is the Workmanship of Men."

subject matter and argument. This circumstance suggests that Locke in the course of writing the *Essay* became more fully acquainted with the vast body of contemporary literature on language and its nature. This suggestion is clearly supported by the recently published accounts of Locke's reading during his years in France between 1675 and 1679, as well as by the list of books he sent home. They include, for instance, the Port-Royal Logic, the Port-Royal *Grammaire Générale et Raisonnée*, and Lancelot's grammars of Spanish, Italian, and French.<sup>8</sup> Furthermore, travel literature was one of Locke's special interests; it has been suggested that such reading seems a rather trivial occupation for a philosopher, but, of course, travel accounts were to him what the laboratory and experiments were to the men in the Royal Society. Travel accounts gave experimental knowledge of human nature and behavior, including language.<sup>9</sup> It is well to remember that the epigraph from Cicero's *De Natura Deorum*, which Locke originally placed on the title page of the *Essay*, occurs in a context that argues about the names of the gods and whether their existence has anything to do with their having names.<sup>10</sup> We may also relevantly recall the motto of the Royal Society; it is "Nullius

in Verba," and is drawn from this passage in Horace's *Epistles* I, i:

Ac ne forte roges quo me duce, quo lare tuter,  
Nullius addictus iurare in verba magistri,  
Quo me cumque rapit tempestas, deferor hospes.

## II

Locke was fifty-six when the first publication appeared under his name while Leibniz' *Dissertatio de Arte Combinatoria* was in print in 1666 when its author was only twenty—and it was not his first publication. This interest in a philosophical language was one of the chief preoccupations of the century from its early decades and was by no means original with Leibniz, though he carried it beyond any stage reached by his predecessors. It bears an intimate relation both to his philosophy at large and to his life-time interest in other forms of language study.<sup>11</sup> This interest covered vernacular dialects both old and new, the German language, etymology, particular vocabularies of many sorts, and the nature of language itself. On these matters Leibniz was almost certainly more widely informed than Locke, though Locke undoubtedly knew much more than is immediately

<sup>8</sup> See Lough, "Locke's Reading," *passim*. Some of these may have diverted Shaftesbury during his captivity in the Tower in 1677; on 16 Aug. 1677, Thomas Stringer wrote to Locke: "He apprehends there are some books both Latin and French, and other Janua linguarum or Colloquies, and also he desires to know what grammars. . . . Having your order, I opened the box of things, and have furnished him with those books you sent over. He has engaged to be very careful in restoring them; and in order thereunto, hath got a box to keep them in, apart from all other things." See King, *Life*, p. 39.

<sup>9</sup> Most of the *Essay*'s infrequent references are to travel literature. Locke, *Works*, vol. IX, pp. 512-554, give "A Catalogue and Character of most Books of Voyages and Travels" in Latin, Spanish, Italian, French, and English; there is some reason to believe that it was compiled by Locke himself. Shortly after his return from France, Locke wrote to Thoynard on 15 July 1679, thanking him for the books he had sent over and asking about their cost, "puisque je souhaite fort un commerce de la sorte fort libre de bons livres qui s'impriment chez vous (et surtout des bonnes relations des voyages) et j'apprends que notre pays en est trop stéril pour me fournir assez de cette sorte de marchandise pour entretenir avec vous cette commerce comme il faut." Henry Ollion, ed., *Lettres inédites de John Locke* (La Haye, 1912), p. 28. In his letters to distant travelers, Locke never failed to include languages among his requests for information; see King, *Life*, p. 25, Locke to C. Cudworth, 27 April 1683, the young man having "got quite to the other side of the world." Locke wrote: "I should trouble you also with inquiries concerning their languages, learning, government, manners," where "languages" is inserted above the line (MS. Locke c. 24, fol. 35). Early in 1679 Locke was reading F. Geb. Sagard Theodat, *Histoire du Canada* (Paris, 1636) and on 3 March notes: "Letters. Hurons have no labials in their tongue." (MS. Locke c. 33, fol. 10). The same memoranda book contains these references: "Acta Eruditorum, Lipsiae 82, Rogue de l'origine de noms & de surnoms 12° Paris 81-85" (fol. 32v), and "Morhofius de Germanorum lingua & poesi 8° Kiel 82-271" (fol. 34). The Journal for 1683 under 17 May has this information drawn from "The English Atlas, vol I fol. Oxford 1680": "The Gronlanders call their king Cachico . . . and Casique being the name of their kings from thence quite to Chile one may imagine them all of one original" (MS. Locke f. 7, p. 100).

<sup>10</sup> The Cicero passage occurs in Book I, xxx (or paragraph 84) and is spoken by Cotta for the academics against the Epicurean doctrines expounded by Velleius in the early part of the book.

<sup>11</sup> John Wilkins' *Essay Towards a Real Character and a Philosophical Language* (1668) was known to Leibniz when he wrote his first letter to Oldenburg on 13/23 July 1670: "A Cl<sup>mo</sup> Hezenthalero habeo, quae de perfecta quadam ad usum Philosophiae linguae condenda celeberrimus vester Wilkensius deliberet." *Akademie Ausgabe*, 2nd ser., vol. I, p. 60. The English resident in Frankfurt, William Curtius, wrote to Leibniz on 13/23 Oct. 1670 that he was sending a copy (*ibid.*, p. 67), and on 29 April/9 May 1671, Leibniz wrote to Oldenburg that he had read it recently (*ibid.*, 104). Locke's emphatic rejection of the viability of a philosophical language (implicitly as an inescapable consequence of his analysis of words and language, explicitly in III, xi, 1-2) contrasts with Leibniz' belief in it. The problems I deal with in this paper are complementary to an understanding of this opposition in regard to the philosophical language. Some readers may later in this paper see intimations of my argument on that point.

evident. Leibniz, however, had written a great deal on language before he came to the writing of his *Nouveaux Essais*, and his position is therefore easier to define.

In 1670 Leibniz edited a work by the Italian humanist Marius Nizolius, called *Anti-Barbarus seu de Veris Principiis et Vera Ratione Philosophandi contra Pseudophilosophos* (1553). For his edition Leibniz wrote a Preface on philosophical style, which is conventional for its time and not very different from many contemporary German works that dealt with the relationship between philosophy and knowledge on the one hand, language and words on the other. It presents some of the principles he also later adhered to:

Illud igitur pro certo habendum est, quicquid terminis popularibus explicari non potest, nisi immediato sensu constet (qualia sunt multa genera colorum, odorum, saporum) esse nullum, et a philosophia velut piaculari quodam carmine arcendum.<sup>12</sup>

This is the same basic tenet as Descartes voiced in his remark about the language of Lower Brittany. Leibniz continues that scholastic obscurity had been abandoned sooner in England and France than in Germany because the English and French have long been accustomed to cultivate philosophy in their own tongue; but with his German contemporaries and predecessors in these concerns, he is at the same time convinced that the German language has a philosophical mission, for it is especially rich in good, concrete terms pertaining to the "artes reales" and thus does not easily allow bad philosophy:

Illud tamen asserere ausim, huic tentamento probatorio atque examine philosophematum per linguam aliquam vivam, nullam esse in Europa linguam Germanica aptiorem, quia Germanica in realibus plenissima est et perfectissima, ad invidiam omnium caeterarum, cum artes reales et inter eas mechanicae a multis seculis a nulla gente sint diligentius excultae, usque adeo ut ipsi Turcae in

fodinis Graeciae et Asiae minoris vocabulis metallicis Germanorum utantur. Contra ad commentitia exprimenda lingua Germanica est facile ineptissima, longe quidem Gallica Italicaque et caeteris Latinae propaginibus ineptior. . . . Sed si terminos Latinos retinere aut detorquere voluisset aliquis, hoc jam erat non germanice sed latine philosophari. . . . Atque haec causa fuit philosophiae apud nos serius vernaculo sermone tractatae, quia Lingua Germanica non a philosophia quidem, a barbara tamen philosophia abhorruit, cum vero barbara philosophandi ratio sero sit pulsa, mirum etiam non est, tardigradam in philosophando nostram linguam fuisse.<sup>13</sup>

Behind these words lies the conviction that a language that is unmixed and original—that has, in other words, retained in its own stream the greatest number of root-words ever since Babel and the Confusion—is somehow in closer accord with the Cosmos and Nature, with Creation and the Creator, with truth and reality. Consequently it is also better able than other languages to record the truths of natural philosophy, for it is itself "natural." German was as original as Hebrew and not inferior to Latin and Greek. Leibniz was devoted to Justus Georg Schottel's great epitome of the study of language, and especially of German, since the Reformation, the *Ausführliche Arbeit von der Teutschen Haubt-Sprache* (Braunschweig, 1663), which had said:

Last uns . . . unser Teutsch in Betracht nehmen/ und uns besinnen/ mit was kräftiger kurtzer Ausrede und lieblichem Geschalle/ nach Geheiss der innersten Eigenschaft/ die Teutsche Sprache sich hören lasset/ und dem Lateinischen und Griechischen in naturmessiger Vorstellung ohnzweifelich neben-wo nicht bevortrit. . . . Wir haltens dafür/ dass man den Namen eines Dings aus dessen Natur und Weise zufoderst erforschen und herführen sol. "Illa lingua," sagt Bibliander [*De ratione communi omnium linguarum et literarum* (1548)] "est omnium perfectissima, cujus verba rerum naturas explanant. Qualem credibile est fuisse illam, quā Adam singulis rebus nomina imposuit."<sup>14</sup>

<sup>12</sup> Gerhardt, vol. IV, p. 143.

<sup>13</sup> Gerhardt, vol. IV, p. 144. The same observations occur in Leibniz' *Unvorgreifliche Gedanken*, § 9-11.

<sup>14</sup> Schottel, pp. 60-61. Cf. p. 61: "Nun hat die mildreiche allgemeine Mutter/ die gütige Natur/ auch dieses sonderlich den Teutschen verliehen/ dass sie durch Behülff der Lippen/ Zungen/ Zähnen und Kehlen/ fast unendlichviele einsilbige Wörter können ausreden/ darunter auch die Stammwörter/ als eines einzigen Dinges einlautende Anzeigen seyn." And further p. 144: "Weil demnach in unser Teutschen Sprache die Wurtzeln fast überall einlautend/ klar/ hell/ deutlich/ vollkommen/ ja solche seyn/ darin die Natur ihr Meisterstück getahn/ und ihre Verborgenheit lassen dadurch aufs gründlichste ausbilden; Auch weil unsere Uralte Vorfahren solche nicht von ausländischen Völkern und ihren Feinden erbettelt/ sondern vielmehr bey Babylon sie erlernet haben/ . . . warum wollen wir denn so unbesonnene Unleute gegen unsere Vorfahren seyn/ und ihnen eine Zunge ohn Sprache zuweignen? Wollen wir selbst wider den Lauf der gemeinen Natur/ und wider das Geheiss der Wahrheit aus schändlicher Frömdgierigkeit unsere wunderreiche/ unsere Kraft- und Saftreiche/ reinlichste/ Uralte Stammwörter/ zu Menglingen/ Fliktstücken/ Brokken/ Unwörtern/ Misgeburten und Betteldreкке machen?" For this subject and its history,

Clearly, this doctrine would lose its foundation if language were conventional rather than natural; further, the doctrine rested on the axiom concerning the ancient source of the German language and was sustained by historical study that attempted to reach back to the origin.

Many of the points that Leibniz had made in the Preface to Nizolius were repeated in a work from the early 1680's, his *Ermahnung an die Teutsche, ihren Verstand und Sprache besser zu üben*, whose title contains a characteristic linking of "Verstand" and "Sprache." But *Die Ermahnung* lay unpublished until 1846, unlike a later and better work, *Die Unvorgreiflichen Gedanken betreffend die Ausübung und Verbesserung der Deutschen Sprache*. It was in part written in the late 1690's, revised and expanded during the next ten years, and first published in 1717 in the *Collectanea Etymologica*, a collection of various pieces—by no means all by Leibniz—that was made by his former secretary Johann Georg Eccard.<sup>15</sup> *Die Unvorgreiflichen Gedanken* exercised a steady and strong influence in Germany throughout the eighteenth century and beyond. But during his own lifetime the only public and international record of his interest in etymology and even in what we may more nearly call philology was the Latin treatise that stood at the head of the first volume of the *Miscellanea Berolinensia*, the memoirs of the Berlin Academy, whose founder Leibniz was. It appeared in 1710 and was the "Brevis designatio meditationum de originibus

gentium, ductis potissimum ex indicio linguarum." In the eighteenth century it influenced a few German scholars—chiefly Johann Georg Wachter—but did not otherwise gain much attention until later, largely owing to that century's bias in favor of universal grammar and the philosophical question of the ultimate origin of language. It opened with a statement that has since become as famous as it was then prophetic:

Cum remotae Gentium Origines Historiam transcendunt, Linguae nobis praestant veterum monumentorum vicem. Et vetustissima linguarum vestigia supersunt in nominibus fluviorum atque sylvarum, quae mutatis accolis plerumque persistunt; proximae sunt locorum ab hominibus constitutorum appellationes; quanquam enim multae villae, multa oppida à conditoribus nominentur, quod in Germania valde est frequens, quae serius exculta est; alia tamen loca à situ, à proventu, à caeteris qualitatibus appellantur, & vetustiorum difficilis est etymologia. Nomina etiam vetera hominum, quorum nulla Germaniae gens plura Frisiis retinuit, ducunt nos in sacraria, ut sic dicam, veteris linguae. Illud enim pro axiomate habeo, omnia nomina quae vocamus propria, aliquando appellativa fuisse, alioqui ratione nulla constarent. Itaque quoties vocabulum fluminis, montis, sylvae, gentis, pagi, opidi, villae, non intelligimus, intelligere debemus, ab antiqua nos lingua discessisse.<sup>16</sup>

The phrase "in sacraria veteris linguae" may stand as a motto for the outburst of philological activity that occurred round 1800.<sup>17</sup> This passage

the best exposition is Wolfgang Kayser's brilliant "Böhmes Natursprachenlehre und ihre Grundlagen," in *Euphorion*, vol. 31 (1930), pp. 521–562, as well as chap. 6 "Die Sprachtheorie als Grundlage für die Verwendung der Klangmalerei," in Kayser's *Die Klangmalerei bei Harsdörffer* (Leipzig, 1932), pp. 137–186. I cannot entirely agree with Kayser's polemic against Paul Hankamer's *Die Sprache, ihr Begriff und ihre Deutung im 16. und 17. Jahrhundert* (Bonn, 1927), to the effect that Böhme's "Natursprachenlehre" has nothing to do with "Klangmalerei" in German Baroque poetry; Kayser says little about Quirinus Kuhlmann, who is the crucial text. See also August Schmarsow, *Leibniz und Schottelius. Die Unvorgreiflichen Gedanken*, untersucht und herausgegeben (Trübner: Strassburg und London, 1877); it is vol. 23 in *Quellen und Forschungen zur Sprach- und Culturgeschichte der Germanischen Völker*. And Friedrich Gundolf, "Justus Georg Schottel," in *Beiträge zur neueren Literaturgeschichte* vol. 16 (1930), pp. 70–86.

<sup>15</sup> For the dating and the textual problems of *Die Ermahnung* and *Die Unvorg. Ged.*, see Paul Pietsch's edition in "Leibniz und die Deutsche Sprache" in *Wissenschaftliche Beihefte zur Zeitschrift des Allgemeinen Deutschen Sprachvereins*, Hefte 29 (July 1907) and 30 (April 1908), pp. 265–371, esp. pp. 290–291 and 313–327.

<sup>16</sup> These points are clearly anticipated in Schottel's *Ausführliche Arbeit* in the Introduction to the long alphabetical register of German proper names given in "Libri V, Tractatus Secundus: De Nominibus Propriis veterum Teutonicorum seu Celticorum populorum." See esp. p. 1034: "Es ist nicht zuleugnen/ dass in den manigfaltigen Nahmen der Stäte/ Berge/ Wälder/ Länder/ Männer/ Weiber/ &c Ungewissheit/ und dunkler Verstand dem/ so der gründlichen Andeutung nachsinnet/ oft vor Augen stehe/ welches ohn zweifel daher mit rühret/ dass viele alt Teutsche Wörter uns in jetziger unsere Mundart unbekant worden/ ein Wörterbuch aber/ darin solthane Wörter sollen verzeichnet und erklärt seyn/ ist nicht verhanden." The axiom that proper names derive from appellatives had been illustrated in the short piece called "Aliquot nomina propria germanorum ad priscam etymologiam restituta," which since 1554 had been ascribed to Luther and consequently exercised a considerable influence. Its first entry is typical: "Rodolfus, a latinis corruptum est, Sed Germanice dicitur 'Rathulff' id est 'consilium et auxilium' salus enim seu auxilium dicitur 'Hulff', 'Rat' 'consilium'." Similar analysis had already been applied to the names that occur in the Old Testament. Though the work is not believed to be Luther's, it is printed in the Weimar edition of *Luthers Werke*, vol. L, pp. 135–159, with an introduction and notes.

<sup>17</sup> See Sigrid von der Schulenburg's excellent monograph "Leibnizens Gedanken und Vorschläge zur Erforschung der Deutschen Mundarten," in *Abhandlungen der Preussischen Akademie der Wissenschaften, Philologisch-historische Klasse*, Nr. 2 (1937), pp. 1–37. On his return from England in 1676, Leibniz brought a copy of John Ray's recent *Collection of English Words*, which he

states the very important principle that proper names must at the outset have been appellatives, for otherwise they would exist without any reason, and Leibniz could not allow that sort of inconsistency in his philosophy.

### III

But the work that more nearly concerns us was written some years before the publication of the "Brevis designatio." This work is Leibniz' critique of Locke's *Essay*, his *Nouveaux Essais sur l'Entendement*, which in its discussion of Book III had contained the same observations. The *Nouveaux Essais*, however, did not appear in print until 1765, when they were published by the notorious and fantastic Rudolf Erich Raspe. This fact is of considerable historical importance, for thus Locke alone came to influence the lively concern with language among the French *philosophes*, spurred by Voltaire's dictum in the *Lettres Philosophiques*: "Tant de raisonneurs ayant fait le roman de l'âme, un sage est venu, qui en a fait modestement l'histoire. Locke a développé à l'homme la raison humaine, comme un excellent Anatomiste explique les ressorts du corps humain." This goes for Condillac, Turgot, and many others as well as for a number of articles in the *Encyclopédie*, including Turgot's influential article on etymology.

In Germany the *Nouveaux Essais* gained a wide and rapid influence. Raspe's edition was noted in the *Göttingische Anzeigen* under 10 Jan. 1765, with special mention of Book III, ii as offering "ein Vorrath gelehrter Sprachkunde und derselben Anwendung auf den Ursprung der Völker, den freylich nur ein Philosoph, wie L[eibniz] war,

darbringen konnte. Dass fast alle Sprachen gemeinschaftliche Wurzelwörter und andre offenbare Uebereinstimmungen haben, scheint L. einen gemeinschaftlichen Ursprung der Menschen darzuthun, gegen den manche die Verschiedenheit der Sprachen als einen Einwurf ansehen." Having suggested a German translation of Monboddo's great work *Of the Origin and Progress of Language*, Herder wrote a Preface to the first volume, which came out at Riga in 1784. He rightly saw that the work was much more assured of success in Germany than in England and gave the revealing reason: "Locke geht uns nicht weiter an, als sofern er der Wahrheit diene, und wir sind lange schon durch Leibnitz gewöhnt, auch schwache Seiten seiner Philosophie zu finden."<sup>18</sup>

I shall not give a detailed account of Leibniz' fruitless effort to enter into direct correspondence with Locke; this history has been set forth in detail in the Introduction to the recent volume of the Leibniz *Akademie Ausgabe* that contains the *Nouveaux Essais* and related pieces. The *Essay* was brought to Leibniz' attention in Dec. 1690, but we have no proof that he became even somewhat acquainted with it until June 1695.<sup>19</sup> His first piece on the *Essay* was "Quelques remarques sur le livre de Mons. Lock intitulé *Essay of Understanding*," which he sent to Burnett in March 1696, who, however, did not communicate it to Locke until 24 March 1697, regretting that he could not "recover it sooner out of the hands of those that were more curious to see than willing to restore it. My own coppie was so much used to pieces that I have sent you a trew double of it."<sup>20</sup> Locke had then already with Cunningham "had a sight of it last summer, and he and I read it paragraph by

sent to Gerhard Meier in Bremen on 16 July 1694 with word that a new edition had just appeared. Later, in 1712, Leibniz sent a still unpublished "Epistolaris de Historia Etymologica Dissertatio" to Eccard with advice, in the words of the monograph, that "wie John Ray die Landschaften seines Vaterlandes bereiste und neben den Pflanzen ihre besonderen Wörter sammelte, so soll Eckhart die deutschen Lande durchwandern und den eigenen Sprachgebrauch eines jeden erforschen, um dann auch die andern Germanischen Ländern heimsuchen" (p. 5). These were the so-called "Landwörter," which in Germany received much attention during the eighteenth century. For additional evidence of Leibniz' preoccupation with language at the time when he was working with the *Essay*, see especially the letters to John Gabriel Sparwenfeld in J. F. Feller, *Optum Hanoveranum* (Leipzig, 1718), pp. 32-39; C. Kortholt, *Leibnitii Epistolae ad Diversos*, vol. III. (Leipzig, 1738), pp. 355-356; C. Vilh. Jacobowsky, J. G. Sparwenfeld, *Bidrag till en Biografi* (Stockholm, 1932), *passim*. The letters in Feller and Kortholt contain much matter that also appears in the *Nouveaux Essais*, Bk. III, and in the "Brevis designatio." The best edition of this correspondence is Harald Wieselgren, "Leibniz' bref till Sparfvensfelt," in *Antiquarisk Tidskrift för Sverige*, vol. 7, (1884-1885), No. 3, pp. 1-64. The ten letters reprinted here date between Dec. 1695 and Jan. 1700 and are of capital importance for an understanding of Leibniz' linguistic doctrines and knowledge; it is evident that Sparwenfeld supplied Leibniz with much information that was later included in the *Nouveaux Essais*.

<sup>18</sup> B. Suphan, ed., *Herders Sämmtliche Werke*, vol. XV, p. 180.

<sup>19</sup> Gerhardt, vol. III, p. 162, in letter to Burnett of 11/21 June 1695.

<sup>20</sup> MS. Locke c. 4, fol. 197. In this as in other letters, Burnett is very clearly trying to lure Locke into answering and giving his opinion. He continues: "I am to wryte back to him againe within 10 or 12 days & iff you have any word to him, or any command for me I shall be glad of an opportunity to serve you both. . . . It is not possible to expresse in a letter the great character Monsieur Leibnitz hath of you as he lately so ampie and often declared Himself to [me]."

paragraph over together, and he confessed to me, that some parts of it he did not understand; and I showed him in others, that Mr. L——'s opinion would not hold, who was perfectly of my mind."<sup>21</sup> But before the arrival of the Burnett copy, Locke had received another, which he mentioned in a letter to Molyneux of 22 Feb. 1696/7; and early in April, he received a third copy via LeClerc, with the latter's comment on Leibniz: "Je croi néanmoins qu'il ne vous entend pas, et je doute qu'il s'entende bien lui-même, ce qui soit dit entre nous."<sup>22</sup> These words closely resemble Locke's well-known and final judgment of Leibniz, expressed in the letter he wrote to Molyneux the next day:

I must confess to you, that Mr. L——'s great name had raised in me an expectation which the sight of his paper did not answer, nor that discourse of his in the "Acta Eruditorum," which he quotes, and I have since read, and had just the same thought of it, when I read it, as I find you have. From whence I only draw this inference, that even great parts will not master any subject without great thinking, and even the largest minds have but narrow swallows.

Leibniz did not see these unkind words until after the publication of *Some Familiar Letters* in 1708, but he had heard nearly the same opinion from Burnett in a letter of 16 Dec. 1701:

Il ne se soucie pas d'avoir dispute plus avec personne à l'égard de ce qu'il a écrit; mais j'ay ouy dire aussi qu'il ne pretendoit pas assez comprendre ce que vous aviez écrit pour vous repondre, ce que je vous écris en ami: mais je ne veux pas que vous me nommiez pour cette notice, n'ayant jamais aucuns ordres de vous écrire autres choses que ses remerciemens et son estime pour vous.<sup>23</sup>

Whether Locke was not up to controversy or criticism in 1701 is one thing; it is another that he strongly encouraged critical comments in his letter to Limborch of 29 Oct. 1697:

Je souhaiterois que mon *Essai* . . . fut écrit dans une langue que ces excellens hommes pussent entendre, car par le jugement exact & sincere qu'ils porteroient de mon ouvrage, je pourrois compter surement sur

ce qu'il y a de vrai ou de faux, & sur ce qu'il peut y avoir de tolerable,

and he goes on to mention with regret that the third edition has not been doing as well as the first two.<sup>24</sup> Thus in spite of Leibniz' repeated attempts, in spite of Burnett's frequent assurances of Leibniz' profound and no doubt sincere respect for Locke, and in spite of Locke's own eagerness to receive criticism from others at the very same time, Leibniz never reached the man, whom of all his contemporaries he was most eager to engage in a philosophical exchange of ideas. A few years later, Leibniz found some comfort that he was indirectly reaching Locke through the correspondence with Lady Masham that began within a year of Locke's death—"il y a apparence qu'il y a quelque part, au moins par le jugement qu'il en faisoit sans doute, et qu'il ne dissimuloit pas apparemment auprès de cette dame," were Leibniz' words to Burnett in August 1704.<sup>25</sup> The consequence of Locke's silence was the extensive discussion of Locke's *Essay* that Leibniz presented in the *Nouveaux Essais*. The work was written between the summers of 1703 and 1705 and was made possible by the appearance of Coste's French translation of the *Essay* in the summer of 1700—Leibniz' English was not good enough to insure a correct reading of Locke's text, a matter on which Locke remarked in a letter to Molyneux on 22 Feb. 1696/7: "I suspect he has, in some places, a little mistaken my sense, which is easy for a stranger, who has (as I think) learned English out of England." Leibniz had planned to publish his discussion, but after Locke's death he decided not to do so because he did not wish to send abroad his arguments against a man who was no longer alive to defend himself.

Why Locke professed not to understand Leibniz is the philosophical question not of one century but of at least two. We cannot, perhaps, dismiss the possibility that Locke may have been influenced by the early rumblings of the undignified priority controversy over the calculus, which had already begun to take shape with Newton's cagey letter to

<sup>21</sup> Locke, *Works*, vol. VIII, p. 407, Locke to Molyneux 10 April 1697.

<sup>22</sup> This copy of "Quelques remarques . . ." is found in MS. Locke c. 19, fol. 162-165, and it is endorsed in Locke's hand: "Leibnitz sent me by M<sup>r</sup> le Clerc 9 Apr. 97." The text of Le Clerc's letter is printed in Gabriel Bonno, ed., *Lettres Intédites de Le Clerc à Locke* (Berkeley, 1959), pp. 98-99.

<sup>23</sup> Gerhardt, vol. III, p. 281.

<sup>24</sup> Locke, *Works*, vol. IX, p. 63; this part of the letter is written in French to make its contents easier to communicate to "ces excellens hommes," who did not read Latin, which is otherwise the language of the Locke-Limborch correspondence.

<sup>25</sup> Gerhardt, vol. III, pp. 297-298. Cf. Lady Masham to Leibniz, 3 June 1704, on Locke's respect for Leibniz, in Gerhardt, vol. III, p. 351.

Leibniz of 16 Oct. 1693, written to cushion some of the potential unpleasant consequences of the recent publication in John Wallis' *Opera* II of some of the letters from the 1670's that were supposed to show Leibniz' dependence on Newton in regard to the calculus.<sup>26</sup> But even without that unpleasant matter I am convinced that Locke had sufficient grounds for not responding to Leibniz. Their opposition was too fundamental.

To provide the fullest and fairest statement of Locke's doctrines, Leibniz wrote his *Nouveaux Essais* as a dialogue between a "lover of Truth" Philalethe and the "lover of God" Theophile; the latter is Leibniz while the former speaks for Locke and corresponds to Pierre Coste.<sup>27</sup> The procedure is simple. Philalethe presents Locke's views by direct quotation, paraphrase, summary, or a mixture of all three at such length as the matter may justify, often with omission of illustrations and examples; then Theophile answers. Sections may be telescoped into a single paragraph or entirely omitted; chapters may be severely shortened but none are completely left out.

Departure from Locke's order occurs occasionally and shows that Leibniz took some care with the composition. For instance, most references to words and language in Book II are omitted and some of them transferred to Book III. A noteworthy example occurs in II, xi, 10-11, where Locke observes that animals "have not the faculty of abstracting, or making general *Ideas*, since they have no use of Words, or any other general Signs." Leibniz transposes most of this discussion to the opening pages of Book III, where it is in fact more appropriate. Another sort of significant change is found in the chapter "Of Enthusiasm," which Locke in the fourth edition (1700) added as IV, xix. It deals with the visions, perceptions, and language of mystics, but Locke gives no names; Leibniz, however, adds the names that must almost certainly have been in Locke's mind: Boehme, Quirinus Kuhlmann, Christine Poniatowa, Drabitus, and their defender Comenius, all made infamous—and Comenius much maligned—by the recent publication of Bayle's *Dictionnaire* (1697).<sup>28</sup> This is only one among many instances

<sup>26</sup> See H. W. Turnbull, ed., *The Correspondence of Isaac Newton*, vol. III (Cambridge, 1961), p. 285, and note the words: "Nam quamvis commercia philosophica & mathematica quammaxime fugiam, tamen metuebam ne amicitia nostra ex silentio decrementum acciperet." On the controversy, see J. E. Hofmann, *Geschichte der Mathematik*, vol. II and vol. III (Berlin, 1957), pp. 75-93 and pp. 4-24. Leibniz' first offense according to the English was his presumed failure to refer to Newton in his "Nova Methodus pro Maximis et Minimis" in the *Acta Eruditorum* for Oct. 1684 (pp. 467-473), which by 70 pages preceded his "Meditationes de Cognitione, Veritate, et Ideis" from November of the same year; this was the item to which Leibniz referred in the piece on the *Essay* that Locke saw in Feb.-Mar. 1697 and of which he expressed his low opinion to Molyneux in the letter of 10 April 1697. It is worth noting that Locke owned the *Acta Eruditorum* and took care to keep up the series (MS. Locke fol. 29, p. 4: "Acta Eruditorum I have 82. 83. 84. 85. 86." and Wetstein's letters to Locke in MS. Locke c. 23, fol. 75 ff.). Oldenburg's supposed role in giving Leibniz access to Newton's material is well known; it is therefore interesting that Leibniz late in the 1690's still did not have the *Philosophical Transactions* that had appeared since Oldenburg's death. MS. Locke c. 13, fol. 156a-156b, contains a long list of items Leibniz hoped to acquire through Burnett, including "14° Je desire d'obtenir d'Angleterre toutes les philosophiques transactions imprimées depuis celles de feu M. Oldenbourg ou plus tost depuis l'index general." In Feb. 1699, Leibniz again made the same request to his amanuensis in London, F. A. Hackman: "Je vous supplie Monsieur, de vous en souvenir pour moy aussi bien que des Transactions depuis la Mort de M. Oldenbourg ou depuis l'index General." See Paul Ritter, "Neun Briefe von Leibniz an Friedrich August Hackman" in *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (1915), p. 724.

<sup>27</sup> Leibniz very carefully introduces the two participants in the dialogue at the beginning of Book I. It is clearly important to keep the two and their opinions apart and not take statements from Philalethe as if Leibniz subscribed to them; yet, surprisingly, this distinction has not always been observed. Gustav Konrad, at the beginning of his *Herders Sprachproblem im Zusammenhang der Geistesgeschichte* (Berlin, 1937), cites Philalethe's rendering of the opening paragraph of Book III on the social function of language as if this doctrine were Leibniz'; even if it is, the matter is not so simple. The historical consequences are of course bizarre, and Konrad does not escape them. The very same confusion occurs in Siegfried Körniger, "G. W. Leibnizens Sprachauffassung," in *Die Sprache, Zeitschrift für Sprachwissenschaft*, vol. 4 (1958), pp. 4-14. In general it may be said that German treatments of Leibniz on language are not what one would have hoped to find; they tend to be vitiated by a desire to make him a sort of Fichte on this subject, a notion that is not at all justified, least of all by *Die Unvorg. Ged.* which is most often cited in this connection. It is true that this work argues for the use of German words if they serve the purposes at hand, but Leibniz also argues for the free adoption of useful foreign terms that have no German counterparts; his attitude is by no means that of the mere purist (see, e.g., *Unvorg. Ged.* § 15-18).

<sup>28</sup> The inception "Of Enthusiasm," however, preceded 1697, as appears from Locke to Molyneux 8 Mar 1694/5 in Locke. *Works*, vol. VIII, pp. 350-351: "I have had some thoughts myself, that it would not be possibly amiss to add, in lib. iv. cap. 18, something about enthusiasm, or to make a chapter of it by itself." Returning from Holland late in 1697, John Wynne brought a copy of Bayle's *Dictionnaire*, which Locke, according to MS. Locke c. 23, fol. 129-132, had earlier commissioned him to buy. Locke was certainly familiar with two of the most notorious enthusiasts. His journal for 1678 contains under "Tuesd. Apr. 19" a brief shorthand entry marked "Drabicus Prophesie" in Locke's own hand; a transcription of this entry is printed in John Lough, *Locke's Travels in France 1675-1679* (Cambridge, 1953), p. 190. All the information, the details, and the names of this

of the light that the *Nouveaux Essais* may throw on the *Essay*. I have only discovered one case in which Locke, or rather Philalethe, is made to concede a point which he would hardly have granted. It occurs in the course of the discussion of the real and nominal essences of substances and mixed modes, in III, x, 21 near the end of the long analysis "Of the Abuse of Words." Here Philalethe is made to say: "Je vois enfin que j'aurois eu tort de blamer ce rapport aux essences et constitutions internes, sous pretexte que ce seroit rendre nos paroles signes d'un rien ou d'un inconnu." At the end of the same chapter occurs one of several instances of humor and good spirit; Locke has just delivered a diatribe against eloquence as tending to insinuate false ideas, to excite the passions, and to seduce reason. Theophile answers that he admires the zeal for truth and does not wholly despair of it "parce qu'il semble, Monsieur, que vous combattés l'Eloquence par ses propres armes, et que vous avés meme d'une autre espece, superieure à cette trompeuse."

## IV

The fundamental opposition between Locke and Leibniz is clearly brought out by a statement Leibniz made in a letter to Burnett as early as 22 Nov. 1695:

Ce n'est pas que ce livre ne soit plus important et plus profond [que le livre de M. Lock de l'Education] et ne contienne quantité de belles choses, mais c'est parcequ'ayant eu moy même le temps et l'occasion d'approfondir peut estre plus que luy cette matiere de l'entendement de l'homme et de l'art de penser, et ayant joint la theorie à la pratique, en donnant plusieurs decouvertes qui ont paru solides, je suis peut estre plus en estat de parler à fonds de la recherche de la verité.<sup>29</sup>

This judgment Leibniz never found occasion to revise; he took the *Essay* to be a "recherche de la vérité." Nearly all eighteenth-century readers made the same mistake about the scope and intent of the *Essay*, which perhaps would not have gained its supreme authority as the arbiter of Truth if it had not been for this mistake. The readers found a pattern they were familiar with and did not stop to think that they had been presented with something quite different. Epistles to the Reader are rarely read with care, nor, even if read, are they taken seriously. The *Essay* had a private origin. It was written "by incoherent Parcels" for "my own Information, and the Satisfaction of a few Friends"; but "those who advised me to publish it, advised me . . . to publish it as it is." Hence "I plainly tell all my Readers, except half a dozen, this Treatise was not at first intended for them; and therefore they need not be at the Trouble to be of that Number." Eustace Budgell said of Robert Boyle: "He has *animated* Philosophy; and put in *Action* what before was little better than a *speculative* Science."<sup>30</sup> Similarly, Locke put the mind and its philosophy in action. Both Boyle and Locke had the conviction that knowledge accumulates by a process that steadily may bring us closer to truth, but the Truth is never achieved; the process never ends. Boyle, Birch tells us, "wanted no capacity or abilities to have worked up a glorious system," but "he nobly despised this poor satisfaction and mean gratification, telling us plainly and expressly, that, notwithstanding all he had done, all the labour, pains, and expence bestowed in a life of natural inquiries, notwithstanding the vastly numerous and important observations and discoveries he made, he saw nothing but the first drawings of science."<sup>31</sup> Huygens and Leibniz, however,

entry are also contained in MS. Locke b. 4, fol. 40-48, which on fol. 47v is endorsed in Locke's hand "Drabicius 25. Apr. 78." This document is a letter in French addressed to Louis XIV, setting forth the writer's own account of his first acquaintance with the prophecies, his initial doubt, and his final conversion to full conviction of their truth; the writer now exhorts the king to listen to the prophet's words and to conduct his policies in accordance with them for the ultimate purpose of securing peace in Europe. Both the autobiographical facts of the letter as well as its historical context make it virtually certain that it is a version of a letter written by the German poet, visionary, and enthusiast Quirinus Kuhlmann, who became so persuaded of the inspired truth in Drabicius' prophecies that he felt bound to carry their message to potentates both in the East and West—or as Bayle has it, "the reading of Drabicius completed the ruin of Kuhlman." He brought—or sought to bring—similar messages to Smyrna, Constantinople, and Moscow. Coming from London, he had passed through Paris shortly before the matter of Drabicius came to Locke's attention, arriving past the middle of March and departing on Palm Sunday, 2 Apr. (See Claus Victor Bock, *Quirinus Kuhlmann als Dichter* (Bern, 1957), p. 45.) It is of course obvious that Locke saw a very close connection between enthusiasm and the "association of ideas." The latter is a "sort of Madness," whose manifestation may be enthusiasm or mysticism, which in the tradition of Boehme depended chiefly on a certain readiness to take words—and especially the words and language of a particular party or sect—for reality. Locke's own words offer a good description of the result: "This gives Sense to *Jargon*, Demonstration to Absurdities, and Consistency to Nonsense" (II, xxxiii, 18).

<sup>29</sup> Gerhardt, vol. III, p. 165.

<sup>30</sup> *Memoirs of the Lives and Characters of the Illustrious Family of the Boyles*, 3rd ed. (London, 1737), p. 120.

<sup>31</sup> Birch, *Life of Boyle*, pp. 317-318.

deplored the absence of what Birch saw as the greatest virtue; Huygens had hoped for "des principes vraisemblables," Leibniz for "une assez grande application pour pousser les consequences autant qu'il faut."<sup>32</sup> In this matter the opposition between the two Englishmen and the two Europeans was complete and explicit. Boyle said in 1666 that he had purposely forborne the perusal of Descartes' system of philosophy,<sup>33</sup> and 20 years later Locke wrote to Clarke that he did not think the works of nature could ever be reduced into a science:

Natural philosophy, as a speculative science, I think, we have none, and perhaps I may think I have reason to say we never shall. . . . Not but that there are many things to be learnt in natural philosophy which abundantly reward the pains of the curious with delight and advantage. But these, I think, are rather to be found amongst such writers as have employed themselves in making rational experiments and observations, than writing speculative systems.

Among such writers Boyle was the best; but if the reader should have "a mind to launch further out into general speculations, I would recommend Descartes' principles, not as perfectly true or satisfactory to an inquisitive man, but yet perhaps the most intelligible and most consistent with itself of any yet to be met with."<sup>34</sup>

Thus one may say that the *Essay* has an esoteric background in the world of Boyle and the men of

the Royal Society. It is not a metaphysical treatise, not a "recherche de la vérité." It does not pretend to offer a complete system of knowledge and truth, but to present a discussion of the ways in which knowledge may be obtained and secured. Its nature is essentially practical, and for that reason it pays much attention to the ways in which we may wrongly come to believe we have certain knowledge when in fact we do not. It is to this problem that Book III "Of Words" is chiefly devoted.<sup>35</sup> It offers many instances of Leibniz' failure to understand Locke's limited scope and intent.

In III, v, 8, "Of the Names of Mixed Modes and Relations," Philalethe has presented Locke's doctrine of intranslatability among languages: "Les hommes formant arbitrairement diverses especes de modes mixtes, cela fait qu'on trouve des mots dans une langue auxquels il n'y a aucun dans une autre langue qui leur reponde."<sup>36</sup> "This," Locke observed in words that Leibniz does not quote, "could not have happened, if these Species were the steady Workmanship of Nature; and not Collections made and abstracted by the Mind, in order to naming, and for the convenience of Communication," which in the previous section he had called "the chief end of Language." Two sections later he reiterates with one of his most telling and characteristic metaphors: "Though therefore it be the Mind that makes the Collection, 'tis the Name which is, as it were, the Knot, that ties them fast together."<sup>37</sup> But to

<sup>32</sup> Gerhardt, ed., *Leibnizens Mathematische Schriften*, vol. II (Berlin, 1849), p. 128 (Huygens to Leibniz, 4 Feb. 1692), and p. 131 (Leibniz to Huygens 9/19 Feb. 1692). Note also Leibniz in the same letter p. 130: "Vous avez raison, Monsieur, de dire que Descartes a parlé d'un ton decisif de l'arrangement des parties de la matiere, cependant ce seroit dommage si nous n'avions pas son systeme. Ainsi je voudrois que Mr. Boyle nous eut laissé ses conjectures."

<sup>33</sup> *Origin of Forms and Qualities* in Boyle, *Works*, vol. III, p. 11. See also Birch *Life of Boyle*, p. 111.

<sup>34</sup> Benjamin Rand, ed., *The Correspondence of John Locke and Edward Clarke* (Oxford, 1927), p. 156. Similar statements occur in *Essay* IV, iii, 26 & 29, and IV, xii, 10.

<sup>35</sup> In his *Diversions of Purley*, vol. I (2nd ed. 1798), p. 31, John Horne Tooke says that Locke's *Essay* has "merely on account of its title, reached to many thousands more than, I fear, it would have done, had he called it (what it is merely) A *Grammatical Essay*, or a *Treatise on Words*, or on *Language*." This view, which is Tooke's starting point, has often been considered plain absurd, but it actually makes better sense than has generally been allowed, for the intervening hundred years had shown a strong tendency to reduce Locke's philosophy to mere sensationalism. On this mistake, see R. I. Aaron, *John Locke*, 2nd ed. (Oxford, 1955), pp. 109-110.

<sup>36</sup> These words are, with insignificant syntactical changes, taken from the Coste version, which is more explicit in its wording than the corresponding passage in the *Essay*. Cf. Locke to Thoynard 29 Nov. 1680: "Je viens de recevoir le Dictionnaire François de Richelet et si j'y emploierais tous les beaux mots qu'il contient, je ne pourrois pas assez exprimer la reconnaissance que j'ay pour votre amitié dont je recoive si grandes marques de tout moment. Il m'estoit rendu après que j'avois écrite ma lettre et le peu que je l'ai consulté il me semble avoir trouvé le vrai secret de faire un bon dictionnaire parce que la manière ordinaire de rendre les paroles d'une langue en ceux d'une autre n'est pas plus raisonnable que d'envoyer quérir un estui en France pour un instrument Anglois dont on ne sçait pas en France ny la forme ny l'usage, parce que les mots des langues différentes ne s'accordent pas mieux que cela." In *Lettres Inédites*, ed. Ollion, p. 83. This work by César-Pierre Richelet is a French-French dictionary with the characteristic title: *Dictionnaire françois, contenant les mots et les choses* (1st ed. Genève, 1680).

<sup>37</sup> This image of the word as a knot that ties a bundle together is one that Locke uses often and with great emphasis; it calls attention both to the arbitrariness of the idea and to the active role performed by the word in preserving the idea as well as in fostering the opinion that it is not arbitrary, on the mistaken assumption that there is some sort of real—and hence not arbitrary—connection between word and object. The relevant passages are: II, xxxii, 6-8 (with the important concluding statement

III, v, 8, Leibniz answers: "La remarque est bonne quant aux noms et quant aux coustumes des hommes, mais elle ne change rien dans les sciences et dans la nature des choses." A few lines later he concludes: "Cependant dans la science même separée de son Histoire ou existence, il n'importe point, si les peuples se sont conformés ou non, à ce que la raison ordonne." I am sure Locke would agree, but he was not talking about "la science" but about "l'histoire." He had stressed this point in I, i, 2, when he said: "I shall imagine I have not wholly mis-employ'd my self in the Thoughts I shall have on this Occasion, if, in this Historical, plain Method, I can give any Account of the Ways, whereby our Understandings come to attain those Notions of Things we have." On 26 June 1681 he had devoted a journal entry to this matter, beginning: "There are two sorts of knowledg in the world generall and particular founded upon two different principles i.e. true Ideas and matter of fact history." He repeated much of that entry in IV, xi, 13-14; Leibniz' long comment on those sections ends in this fashion:

Ces veritez necessaires etant anterieures aux Existences des Estres contingens, il faut bien qu'elles soyent fondées dans l'existence d'une substance neccessaire. C'est là où je trouve l'original des idées et des veritez qui sont gravées dans nos ames, non pas en forme de propositions, mais comme des sources dont l'application et les occasions feront naître des enonciations actuelles.<sup>28</sup>

It is no wonder that Locke did not understand Leibniz; his head would have reeled if he could

have heard this sort of answer to his III, iii, 19: "Je vous ay déjà dit, Monsieur, que les Essences sont perpetuelles parce qu'il ne s'y agit que du possible."

But then, how was Leibniz to know what Locke and his half dozen friends would have taken for granted, the "historical, plain method" of the natural philosophers in the Royal Society? The word "understanding" in the title is to be taken in its active sense as meaning something like "ways of getting to know something"—"mind" is no synonym for it. The *Essay* is about process, not about the still center and the possible. The aim of the *Essay* was much more limited than Leibniz saw; in I, i, 6 Locke had pointed out that "our Business here is not to know all Things, but those which concern our Conduct," i.e., conduct in regard to nature as the Royal Society would have understood it. The design of the *Essay* lay in a "little compass," as Locke wrote to Anthony Collins on 21 March 1703/4: "You have a comprehensive knowledg of it, & do not stick in the incidents, w<sup>ch</sup> I find many people do, w<sup>ch</sup> whether true or false make nothing to the main design of the *Essay*, w<sup>ch</sup> lies in a little compass & yet I hope may be of great use to those who see & follow that plain & easy method of nature to carry them y<sup>e</sup> shortest & clearest way to knowledg."<sup>29</sup>

Locke shared the outlook of his fellow members in the Royal Society. His natural inclinations, his intellectual habits, and his methods were also theirs, though he modestly placed his own efforts below those of a Boyle, a Sydenham, a Huygens, or a Newton. For, as he explained in the "Epistle to

about the "double conformity" between thing-idea and idea-name); III, iii, 20; III, v, 4 & 10 & 13. With these should be taken II, xxii, 2-6, which without using the knot-and-bundle image makes the same point. It is characteristic that Leibniz either runs very lightly over or entirely omits these sections, and thus Locke's image nowhere occurs in his text. III, v, 13 in the Coste translation does not have the image, but the same point is made with equal force.

<sup>28</sup> Cf. also *Nouveaux Essais* III, iii, 14 and IV, v, 2 where Theophile gives some very revealing and typical answers.

<sup>29</sup> Brit. Mus. MS Birch 4290, fol. 38v. This letter was first printed in [P. Des Maizeaux] *Collection of Several Pieces of Mr. John Locke* (London, 1720). It is also in Locke, *Works*, vol. IX, p. 285. Locke offers an excellent gloss on his reasons for limiting his "business" within "a little compass" to those things "which concern our conduct." It is the long entry in the Journal for 8 Feb. 1677, reprinted in Aaron & Gibb, pp. 84-90, especially this passage on p. 88: "The businesse of men being to be happy in this world by the enjoyment of the things of nature subservient to life health ease and pleasure . . . we need noe other knowledg for the attainment of those ends but of the history and observation of the effects and operations of naturall bodys within our power, and of our dutys in the management of our owne actions as far as they depend on our wills i.e. as far also as they are in our power." Therefore "we have noe reason to complain if we meet with difficultys in other things which put our reasons to a non plus confound our understandings and leave us perfectly in the darke under the sense of our owne weakness, for those relating not to our happynesse any way are noe part of our businesse and therefor tis not to be wondered if we have not abilitys given us to deale with things that are not to our purpose, nor conformable to our state or end." This entry echoes a large number of passages in Boyle, e.g., this one written ten years later: "It is true, that to inquire to what purpose nature would have such or such effects produced, is a curiosity worthy of a rational creature, upon the score of his being so: but this is not the proper task of a naturalist, whose work, as he is such, is not so much to discover, why; as how, particular effects are produced." *A Disquisition about the Final Causes of Natural Things*, in Boyle, *Works*, vol. V, p. 443. On these grounds I do not find that the *Essay* and the *Two Treatises* are disparate, though they are not, of course, systematically similar treatments of different subjects, but both deal with our "conduct."

the Reader," "'tis Ambition enough to be employed as an Under-Labourer in clearing the Ground a little, and removing some of the Rubbish that lies in the way of Knowledge." This rubbish he immediately proceeded to identify. "Philosophy, which is nothing but the true Knowledge of Things," had been impeded by the mistaken trust in language and words.

Vague and insignificant Forms of Speech, and Abuse of Language, have so long passed for Mysteries of Science . . . that it will not be easie to perswade, either those who speak, or those who hear them, that they are but the Covers of Ignorance, and Hinderance of true Knowledge. To break in upon the Sanctuary of Vanity and Ignorance, will be, I suppose, some Service to Humane Understanding: Though so few are apt to think, they deceive or are deceived in the Use of Words; or that the Language of the Sect they are of, has any Faults in it, which ought to be examined or corrected, that I hope I shall be pardon'd, if I have in the third Book dwelt long on this Subject; and endeavoured to make it so plain, that neither the Inveterateness of the Mischief, nor the Prevalency of the Fashion, shall be any Excuse for those, who will not take Care about the Meaning of their own Words, and will not suffer the Significancy of their Expressions to be enquired into.

This passage, so prominently placed, leaves no doubt about the great importance that Locke attributed to Book III and its subject. It was also a subject that Boyle—though he devoted no separate work to it—constantly returned to, and naturally so, for what can a chemist accomplish if he starts

with a fixed notion of the real essence of gold. Locke's journals and book lists show that he took care to own Boyle's publications.<sup>40</sup> There can be no doubt that of all men, both living and dead, Boyle was the one whose work and thought Locke knew best and most admired, not least in regard to language and words.

In the *Origin and Forms of Qualities*, first published in 1666, Boyle had said:

It was not at random that I spoke, when in the foregoing notes, about the origin of qualities, I intimated that it was very much by a kind of tacit agreement that men had distinguished the species of bodies, and that those distinctions were more arbitrary than we are wont to be aware of; for I confess that I have not yet, either in *Aristotle* or any other writer, met with any genuine and sufficient diagnostic and boundary for the discriminating and limiting the species of things; or to speak more plainly, I have not found that any naturalist has laid down a determinate number and sort of qualities or other attributes, which is sufficient and necessary to constitute all portions of matter endowed with them, distinct kinds of natural bodies: and therefore I observe that most commonly men look upon these as distinct species of bodies that have had the luck to have distinct names found out for them, though perhaps divers of them differ much less from one another than other bodies, which (because they have been huddled up under one name) have been looked upon as but one sort of bodies.<sup>41</sup>

The implications are sufficiently obvious to go without any other comment than Locke's more general statement of the same principle in III, ii, 5:

<sup>40</sup> MS. Locke, fol. 16, pp. 39, 41, and 38, contain at least 37 items, many of them including several titles. Boyle's writings also frequently appear in the notes concerning books lent and received back or left in boxes with friends and later returned. Locke saw Boyle often during the years between his return from France and his departure for Holland. Locke's earlier close association with Boyle during the Oxford years is a familiar matter.

<sup>41</sup> Boyle, *Works*, vol. III, pp. 49–50. Cf. this passage from the *Sceptical Chymist*: "I might tell you divers examples I have met with, of the contrariety of bodies, which according to the chymists must be huddled up together under one denomination; I leave you to judge, whether such a multitude of substances, as may agree in these slight qualities, and yet disagree in others more considerable, are more worthy to be called by the names of a principle (which ought to be pure and homogeneous) than to have appellations given them, that may make them differ, in name too, from the bodies from which they so widely differ in nature." Boyle, *Works*, vol. I, pp. 532–533. Some of the terms in the passage from the *Origin of Forms and Qualities* are also used by Locke. Cf. "tacit agreement" and Locke in III, ii, 8: "'Tis true, common use, by a tacit Consent, appropriates certain Sounds to certain Ideas." But Boyle's plain-speaking use of "sort" and "kinds" as synonyms for "species" is especially noteworthy, for this usage was also Locke's habit as early as Draft A: "The minde or the man . . . ranks [a certaine number of those simple Ideas] together or else finds them ranked together by others under one general name, which we call a species and if more comprehensive a genus or in plaine English a sort or kinde." Aaron & Gibb, pp. 7–8. Cf. *Essay* III, iii, 15: "The Essence of each Genus, or Sort, comes to be nothing but that abstract Idea, which the General, or Sortal (if I may have leave so to call it from Sort, as I do General from Genus,) Name stands for." In III, vi, 35 "sort" is used as a synonym for "species": "Men make sorts of Things. For it being different Essences alone, that make different Species, 'tis plain, that they who make those abstract Ideas, which are the nominal Essences, do thereby make the Species, or Sort." Again III, vi, 37: "The boundaries of the Species, whereby Men sort them, are made by Men. . . . So that we may truly say, such a manner of sorting of Things, is the Workmanship of Men." The knot-and-bundle is Locke's metaphor for this sorting according to his repeated insistence that "the making of Species and Genera is in order to general Names" (III, vi, 39). Further specific evidence of Locke's reading of Boyle may also occur in his references to the Dasypodius clock in the Cathedral at Strassbourg in III, vi, 3 & 9; cf. Boyle, *Works*, vol II, p. 39; vol. V, p. 169, and vol. V, p. 404. It is also mentioned in Evelyn's *Diary* for June 1645, and in Descartes' letter to Mersenne 8 Oct. 1629.

Because *Men* would not be thought to talk *barely* of their own Imaginations, but of Things as really they are; therefore they *often suppose their Words to stand also for the reality of Things*.

But comforting as this supposition may be, it is false; words bear no reliable relationship to things, being no better than "those *Ideas* we have in our Minds," so that "it is a perverting the use of Words, and brings unavoidable Obscurity and Confusion, whenever we make them stand for anything" else. In the same work, Boyle suggests a historical reason why words can have no authority in regard to things and nature:

For my part that which I am solicitous about is, that what nature hath made things to be in themselves, not what, logician or metaphysician will call them in the terms of his art; it being much fitter in my judgment to alter words [by which he means not the phonetic form of the word but the bundle of ideas a word ties together], that they may better fit the nature of things, than to affix a wrong nature to things that they may be accommodated to forms or words that were probably devised, when the things themselves were not known or well understood, if at all thought on.<sup>43</sup>

This important doctrine is also maintained by Locke. In III, vi, 25, he says that even if "the *real Essences* of Substances were discoverable," we cannot "reasonably think, that the *ranking of things under general Names, was regulated by those internal real Constitutions, or any thing else but their obvious appearances*." He continues:

Since Languages, in all Countries, have been established long before Sciences. So that they have not been Philosophers, or Logicians, or such who have troubled themselves about *Forms and Essences*, that have made the general Names, that are in use amongst the several Nations of Men: But those, more or less comprehensive terms, have for the most part, in all Languages, received their Birth and Signification,

<sup>43</sup> Boyle, *Works*, vol. III, p. 41.

<sup>44</sup> This doctrine is also in Bacon, *Advancement of Learning* Bk. II: "Let us consider the false appearances that are imposed upon us by words, which are framed and applied according to the conceit and capacities of the vulgar sort." *Works*, eds. Spedding, Ellis, Heath, vol. III (London, 1870), p. 396.

<sup>45</sup> See also Boyle, *Works*, vol. III, pp. 27-28 (in *Origin and Forms*). The same observations are made by Jacques Rohault in his *Traité de Physique*, 1st ed. (Paris, 1671), especially in § 2 of the Preface and in pt. I, chap. 4 "Avis touchant les mots." Locke took notes from the *Traité* on 4 March 1678, and late in June he packed it in a box to be sent back to England. Lough, "Locke's Reading," pp. 243 and 248. Cf. Aaron & Gibb, p. 108.

<sup>46</sup> J. Milton French, ed., *The Life Records of John Milton*, vol. IV (New Brunswick, Rutgers University Press, 1956), p. 93. It is not clear whether Boyle himself used this motto; its first appearance in this context would seem to be in the title-page vignette executed by Bourguignon (1699-1773) and used for the Birch edition. (See R. E. W. Maddison, "The Portraiture of the Honourable Robert Boyle, F. R. S," in *Annals of Science*, vol. 15 (1959), p. 158 and Plate VI.) But Boyle often made the same point, as for instance in his will when he wished the Royal Society "a happy success in their laudable attempts to discover the true nature of the works of God," and prayed "that they and all other searchers into physical truths, may cordially refer their attainments to the glory of the great Author of Nature, and to the comfort of mankind." Birch, *Life of Boyle*, p. 292.

<sup>47</sup> II, xi, 17; cf. for the same image and the role of names I, ii, 15.

from ignorant and illiterate People, who sorted and denominated Things, by those sensible Qualities they found in them.<sup>43</sup>

This, of course, explains the distinction between the real and nominal essence of a substance, of which only the latter can be known. Words are no better than men's attention to "obvious appearances" and "sensible qualities." And anyone who thinks otherwise will never become a chemist but remain an alchemist.<sup>44</sup>

## V

Thus the *Essay* was, as it were, intended as a manual in the epistemology of the Royal Society, whose aim was the promotion of natural knowledge. The motive is clearly brought out in Oldenburg's words to Milton in 1656: "Naturam ejusque authorem propius velim contemplari." The motto that appears on the title pages of Thomas Birch's editions of Boyle's *Works* says much the same: "Ex rerum causis supremam noscere causam."<sup>45</sup> Nature or Creation, "the unsealed book of God," revealed the Creator. To insure the success of this august enterprise nothing must be allowed to interpose itself between man and nature, neither innate notions nor words, for either would give man less than a clear view of his object. Locke stressed this objectivity in a familiar passage:

I pretend not to teach, but to enquire. . . . For, methinks, the *Understanding* is not much unlike a Closet wholly shut from light, with only some little opening left, to let in external visible Resemblances, or *Ideas* of things without; would the Pictures coming into such a dark Room but stay there, and lie so orderly as to be found upon occasion, it would very much resemble the Understanding of a Man, in reference to all Objects of sight, and the *Ideas* of them.<sup>46</sup>

Knowledge grows by accumulation, and language is a tool to sort and denominate what comes to be compacted in that box—though often a blunt and inadequate tool. Thus when Theophile at one point says “cependant les idées ne dependent point des noms,” he is missing the mark. Locke does not say that ideas depend on words, but that in practice they generally do; and that we need to realize to avoid error and to increase the order in the box.<sup>47</sup> The *Nouveaux Essais* do not take up the camera metaphor of the understanding, but in the Preface Leibniz uses another that holds a corresponding position in his conception of the mind. This is the equally familiar and in this case also traditional image of the block of marble whose veins outline a statue of Hercules; it is also, characteristically, a static image as any image that is designed to illustrate innateness is bound to be:

Hercule y seroit comme inné en quelque façon, quoyqu'il fallût du travail pour découvrir ces veines. . . . C'est ainsi que les idées et les vérités nous sont innées, comme des inclinations, des dispositions, des habitudes ou des virtualités naturelles, et non pas comme des actions.

But to Locke, the work is the knowledge, not just the means of revelation. No pre-established harmony has traced veins in his understanding or insinuated words that would do the job of veins.

Here Leibniz seems to be involved in an inconsistency, for how can he insist on “cette admirable harmonie préétablie de l'ame et du corps,” while at the same time maintaining that “les idées ne dependent point des noms”? There must somehow be a connection, for language with its words must have its place within the great harmony. The resolution of this difficulty brings us to the fundamental difference between Leibniz and Locke on language, and it also defines their separate allegiances to existing doctrines concerning the nature of language. Strictly speaking, ideas may not depend on words, nor does the reverse hold. The answer is that they are interdependent, thus showing forth one aspect of the total harmony—as suggested in a phrase that occurs in Leibniz' long answer to III, ii, i: “Il y a quelque chose de naturel dans l'origine des mots.” This relationship

may not be apparent from a synchronic view of speech forms such as Locke takes, but it will emerge from an historical and etymological consideration of many languages and dialects. A diachronic view will reveal that these seem to converge toward a common point as we trace them backward in time. Here Leibniz' wide philological knowledge of a great variety of early dialects came to his aid. His historical studies had made him very familiar with the increasing similarity of the German dialects and even of the Germanic languages as illustrated in the Middle Ages. It is this observation that supported his faith that there ultimately is a single radical and primitive language that underlies all the languages of mankind—and he believed this held for the entire earth, for the European languages and the Asian, African, Mexican, Chinese, and American all taken together. Thus if languages do not now reveal the harmony at mere inspection, it will be found to hold all the same.

In this perspective the doctrine of the Adamic language, which in its original purity was lost in the Fall, still retains some of its meaning, though Leibniz did not take that doctrine literally in the manner of, for instance, Paracelsus, Boehme, or Comenius. To Leibniz, as to Schottel and his predecessors, language in the process of time shows a descent from wisdom to a low point that occurred before the “discovery” of this relationship between language and creation, a discovery that in Germany had coincided with the Reformation. To Luther, the Gift of Tongues had given renewed assurance of the Truth contained in this relationship, and thus in his exegesis he replaced the allegorical mode of the Fathers with the Protestant study of the truth that lay concealed in the word, in languages.<sup>48</sup> Adam's naming of the animals was a creative act; it was both a striking manifestation of man's place in creation, of the harmony of the macrocosm and the microcosm, and at the same time a sign of man's control, for in the names he “knew” the creatures. Thus Adam was the greatest of philosophers even after the Fall, as Luther maintains in his lecture on Genesis 2:18–20.<sup>49</sup> Locke's old school-mate Robert South voiced the same doctrine in a sermon on Genesis

<sup>47</sup> Leibniz' words are the answer to Locke's II, xxii, 4 on mixed modes.

<sup>48</sup> See Peter Meinhold, *Luthers Sprachphilosophie* (Berlin, 1958), especially pp. 32–36 and 42–44.

<sup>49</sup> “Quantum mare cognitionis et sapientiae in hoc uno homine fuit! Etsi autem de hac cognitione quoque multum amisit Adam per peccatum, tamen credo, quicquid adhuc in omnium sapientum libris est, qui tot seculis ab eo tempore, quo literae primum natae sunt, scripserunt, id totum eam sapientiam nondum potuit equare, quae tamen postea in Adam haesit, sed paulatim in posteris obscurata, et pene extincta est.” *Werke*, vol. XLII (Weimar, 1911), pp. 90–91.

1:27, preached in St. Paul's on 9 Nov. 1662. Adam, he said,

came into the world a philosopher, which sufficiently appeared by his writing the nature of things upon their names; he could view essences in themselves, and read forms without the comment of their respective properties. . . . An Aristotle was but the rubbish of an Adam, and Athens but the rudiments of Paradise.<sup>50</sup>

Further, since Adam's claim to being both the first and the greatest of philosophers rested on the act of naming, so he also was the first and the greatest of etymologists. In his "Discours sur la Science des Etymologies," first printed in the second edition (1694) of Gilles Ménage's *Dictionnaire Etymologique, ou Origines de la Langue Française*, Beaugrand said: "On peut dire que le premier Homme a esté en mesme temps le premier Etymologiste; & que cette science fut, pour ainsi dire, sa première occupation." It was on these grounds that reformed man in languages and etymology found new guidelines to the reattainment of early knowledge and the wisdom that had been lost.

The mystics and visionaries of the seventeenth century were profoundly influenced by these doctrines, which in turn were strengthened by the mysticism of the Kabbala, which spread over Europe after the expulsion of the Jews from Spain in 1492. The mystic's path to knowledge was the same as Adam's, as Boehme explained in his *Aurora, oder Morgenröthe im Anfang*:

Denn als Adam erstlich geredet hat, so hat er allen Creaturen, nach ihren Qualitäten und instehenden Wirkungen, den Namen gegeben: Und ist eben die Sprache der ganzen Natur, aber es kan sie nicht ein ieder: denn es ist ein Geheimniss, *Mysterium*, welches mir von Gnaden Gottes ist mitgetheilet worden von dem Geiste, der Lust zu mir hat.<sup>51</sup>

Boehme's doctrines were widely known in England where all his works appeared in translation between 1644 and 1663, many of them before the publication of the German originals.<sup>52</sup> And in

England they also received very forceful and complete, though compact, statement in John Webster's *Academiarum Examen* from 1654, which presented a severe critique of the universities, their learning, and their teaching.

In this work Webster, like the men already active in the proto Royal Society, cited Bacon with high admiration; advocated the wider use of English and the mother tongue in order that "arts and sciences may be taught in it, that thereby a more easie and short way may be had to the attaining all sorts of knowledge";<sup>53</sup> and called "this *School Philosophy* . . . altogether void of true and infallible demonstration, observation, and experiment."<sup>54</sup> But unlike those natural philosophers, he also found truth in Plato, Descartes, Sir Kenelm Digby, Robert Fludd, and the Rosicrucians; encouraged the active study of astrology, alchemy, magic, and "Cabalistick Science";<sup>55</sup> and asserted as fundamental doctrine that:

. . . the mind receiveth but one single and simple image of every thing, which is expressed in all by the same motions of the spirits, and doubtlesly in every creature hath radically, and naturally the same sympathy in voice, in sound, but men not understanding these immediate sounds of the soul, and the true *Schematism* of the internal notions impressed, and delineated in the several sounds, have instituted, and imposed others, that do not altogether concord, and agree to the innate notions, and so no care is taken for the recovery and restauration of the Catholique language in which lies hid all the rich treasury of natures admirable and excellent secrets.<sup>56</sup>

He further linked his adherence to this doctrine with profound respect for Boehme and the mystical belief in the reality of the language of nature, for, as Webster said:

I cannot (howsoever fabulous, impossible, or ridiculous it may be accounted of some) passe over with silence, or neglect that signal and wonderful secret (so often mentioned by the mysterious and divinely-inspired *Teutonick* [who of course is Boehme] and in some manner acknowledged and owned by the highly-illuminated fraternity of the *Rosie Croisse*) of the

<sup>50</sup> *Sermons Preached on Several Occasions*, vol. I (Oxford, 1823), pp. 37-38.

<sup>51</sup> *Aurora*, chap. 20, § 91 in *Sämtliche Schriften* (Faksimile-Neudruck der Ausgabe von 1730, ed. Will-Erich Peuckert), vol. I (Stuttgart, 1955), p. 296. The passage is immediately followed by a reference to Boehme's *Mysterium Magnum*, chap. 19, § 22, which reads: "Dass nun Adam in Gottes, und nicht im thierischen Bilde gestanden sey, findet man an deme, dass er aller Creaturen Eigenschaft gewusst hat, und hat allen Creaturen Namen gegeben aus ihrer Essentz, Form und Eigenschaft, er hat die Natur-Sprache verstanden, als das geoffenbarte und geformte Wort in aller Essentz, denn daraus ist ieder Creatur der Name entstanden." *Op. cit.*, vol. XVII, p. 130.

<sup>52</sup> See M. L. Bailey, *Milton and Jakob Boehme: A Study of German Mysticism in Seventeenth-Century England* (New York, 1914), pp. 185-186 and Serge Hutin, *Les Disciples Anglais de Jacob Boehme* (Paris, 1960), pp. 37-42.

<sup>53</sup> Webster, pp. 98-99.

<sup>54</sup> *Ibid.*, p. 68.

<sup>55</sup> *Ibid.*, p. 75.

<sup>56</sup> *Ibid.*, p. 32.

language of nature: but out of profound and deep consideration; must adumbrate some of those reasons, which persuasively draw my judgment to credit the possibility thereof.<sup>57</sup>

Nowhere in the seventeenth century are we closer to what Locke was aiming at when in Book I he rejected what he in his own title of that book expressly called "innate notions." In terms that are highly reminiscent of Boehme, Webster subsequently argues that there must be a language of nature and that this was the language spoken by Adam when naming every living creature:

I cannot but conceive that *Adam* did understand both their internal and external signatures, and that the imposition of their names was adaequately agreeing with their natures: otherwise it could not univocally and truly be said to be their names, whereby he distinguished them.<sup>58</sup>

Thus whatever agreement Webster might have gained among the natural philosophers, he forfeited entirely by his belief in the cognitive value of "innate notions" and by his closely related advocacy of the language of nature.<sup>59</sup> Webster's mysticism violated the objectivity that was the very foundation of natural philosophy, as the Royal Society saw it, and its successful pursuit of certain knowledge of Creation.

Boyle was fully aware of this problem; he therefore firmly rejected the doctrine of the language of nature or, which is the same thing, the language of Adam. In the *Excellency of Theology*, written in 1665, he observed that considering the dim knowledge we are capable of in our present condition, the more reason we have to value religion, since it holds forth hope that we may

attain a condition in which we shall gain "degrees and kinds of knowledge, to which we are here but strangers." He continued:

I favour of which I will not urge the received opinion of divines, that before the fall (which yet is a less noble condition than is reserved for us in heaven), *Adam's* knowledge was such, that he was able at first sight of them, to give each of the beasts a name expressive of its nature; because that, in spite of some skill (which my curiosity for divinity, not philosophy, gave me) in the holy tongue, I could never find, that the Hebrew names of animals, mentioned in the beginning of *Genesis*, argued a (much) clearer insight into their natures, than did the names of the same or some other animals in Greek, or other languages: wherefore, as I said, I will not urge *Adam's* knowledge in paradise for that of the saints in heaven, though the notice he took of *Eve* at his first seeing her, (if it were not conveyed to him by secret revelation) may be far more probably urged, than his naming of the beasts.<sup>60</sup>

That passage may also serve as a reminder of the reason why Locke, Boyle, Newton, and other natural philosophers studied Hebrew and corresponded together about scriptural interpretation; it is only in a later and very ignorant perspective that this occupation has come to look as if it were the business of their idle hours. They had two books to read, Scripture and Nature. Boyle was assured that he might "safely compare several things in the books we call the scripture, to several other in that of nature" and averred that in the perusal of "true naturalists . . . I receive more pleasure and satisfaction, and am induced more to admire the works of nature, than by all their [the poets'] romantick and superficial narratives."<sup>61</sup> Boyle and

<sup>57</sup> *Ibid.*, p. 26.

<sup>58</sup> *Ibid.*, p. 29.

<sup>59</sup> It was precisely in these terms that Webster's attack was answered by John Wilkins (who wrote the Preface) and Seth Ward in their defense from the same year, *Vindicae Academicarum*. Wilkins found that "above all, the man doth give me the freest prospect of his depth and braine, in that canting Discourse about the language of nature, wherein he doth assent unto the highly illuminated fraternity of the *Rosicrucians*. In his large encomiums upon *Jacob Behem*, in that reverence which he professes to judicall Astrologie, which may sufficiently convince what a kind of credulous fanaticke Reformer he is like to prove." Ward, in similar fashion, denied that "ever there was any such Language of Nature" (p. 22).

<sup>60</sup> Boyle, *Works*, vol. IV, pp. 45-46. The naming of Eve is also advanced by Webster: "Now if it be denyed that he understood by his intrinsick and innate light, what she was, and from whence she was taken (which I hold altogether untrue) and that God by extrinsick information told *Adam* from whence she was taken, yet did he immediately give unto her an adaequate name, suiting her original, which most significantly did manifest what was her nature, and from whence it came, and doubtless the name being exactly conformable, and configurate to the *Idaea* in his mind, the very prolation, and sound of the word, contained in it the vive expression of the thing, and so in verity was nothing else but that pure language of nature, which he then spake, and understood, and afterwards so miserably lost and defaced" (*Acad. Ex.*, p. 30). Boyle may well have had Webster in mind, for whereas the latter had cited the "Seraphical Apostle" from I Cor. 13:1 on "the tongues of men and Angels" in favor of his doctrine, Boyle cites I Cor. 13:12 in favor of his: "For now we see through a glass darkly, but then face to face: now I know in part, but then shall I know even as also I am known." *Acad. Ex.*, p. 27 and Boyle, *Works*, vol. IV, p. 46.

<sup>61</sup> Boyle, *Works*, vol. IV, p. 18 (in the *Excellency of Theology*). Cf. *ibid.*, p. 47: "And as there is no doubt to be made, but that, when *David* invented (as the scripture intimates, that he did) new instruments of musick, there was nothing in that invention, that pleased him so much, as that they could assist him to praise God the more melodiously; so the pious student of nature

his eminent friends did not depreciate poetry because they were illiterate and uncultured boors, for obviously if man can hope to gain real knowledge, poetry is of no consequence. The angels do not entertain *themselves* with poetry, though in the human perspective their entertainment is poetry and beautiful music; but for them it is enough to intuit directly. The order they contemplate demands no embellishments.<sup>62</sup>

With his clear awareness of the implications of the doctrine of natural language and "innate notions," we should not expect Locke to fail to make reference to Adam; and indeed he did. The last nine sections of III, vi, "Of the Names of Substances," are devoted to Adam's use of language. The details of the argument are too many to trace, but the "Conclusion" presents this succinct summary:

What liberty *Adam* had at first to make any complex *Ideas* of mixed *Modes*, by no other Pattern, but by his own Thoughts, the same have all Men ever since had.

And the same necessity of conforming his *Ideas* of Substances to Things without him, as to *Archetypes* made by Nature, that *Adam* was under, if he would not wilfully impose upon himself, the same are all Men ever since under too. The same Liberty also, that *Adam* had of affixing any new Name to any *Idea*; the same has any one still, (especially the beginners of Languages, if we can imagine any such,) but only with this difference, that in Places, where Men in Society have already established a Language amongst them, the signification of Words are very warily and sparingly to be alter'd. . . . But in Communication with others, it is necessary, that we conform the *Ideas* we make the Vulgar Words of any Language stand for, to their known proper Significations, (which I have explain'd at large already,) or else to make known that new Signification, we apply them to.

Few chapters in the entire *Essay* show greater and more evident and direct influence from Boyle than this one. Leibniz, however, does not comment on these sections.<sup>63</sup> Locke's long discussion of Adam's language also shows the close texture of his

finds nothing more welcome in the discoveries he makes of her wonders, than the rises and helps they may afford him, the more worthily to celebrate and glorify the divine attributes adumbrated in the creatures." Readers may wish to consult the beginning of Milton's *Of Education*, which in the big edition of the *Prose Works* is accompanied by an introduction and notes that show a grotesque, though not uncommon, miscomprehension of the problem.

<sup>62</sup> Cf. the words that Locke intended to add to III, x, 13: "We cannot but think that angels of all kinds much exceed us in knowledge, and possibly we are apt sometimes to envy them that advantage, or at least to repine that we do not partake with them in a greater share of it. Whoever thinks of the elevation of their knowledge above ours, cannot imagine it lies in a playing with words, but in the contemplation of things, and having true notions about them, a perception of their habitudes and relations one to another. If this be so, methinks we should be ambitious to come, in this part, which is a great deal in our power, as near them as we can; we should cast off all the artifice and fallacy of words." King, *Life*, p. 362. The question of the language of the angels was a commonplace during the seventeenth century. See, e.g., the opening of Samuel Tilly's letter to Locke on 11 Sept. 1655: "To be able to reveal their thoughts by only willing their discourse, is (they say) the privilege of the Angels." In MS. Locke c. 20, fol. 175.

<sup>63</sup> Examples of Locke's similarity to Boyle in this chapter have already been cited in the notes above. Two more points are worth noting. (1) In these sections, III, vi, 43-51, Locke argues against Adam's special authority in regard to language (and hence also in regard to wisdom) in a manner that closely parallels his arguments against Filmer's claims for Adam's parenthood and dominion in the first of the *Two Treatises of Government*. In both places Locke asserts the principle of Adam's ordinary humanity—"and so *Adam's* Sovereignty built on *property*, or as our A[uthor] calls it, *Private Dominion*, comes to nothing. Every Man had a right to the Creatures, by the same Title *Adam* had, viz. by the right every one had to take care of, and provide for their Subsistence; and thus Men had a right in common, *Adam's* Children in common with him." Par. 87 in *Two Treatises*, ed. Peter Laslett (Cambridge, 1960). Locke several times supports his argument with close philological attention to the original Hebrew, e.g. paragraphs 25-27; and at least once, in par. 46, he presents an argument that clearly rejects the uniqueness of Adam's language: "When he [God] vouchsafes to speak to Men, I do not think, he speaks differently from them, in crossing the Rules of language in use amongst them. This would not be to condescend to their Capacities . . . but to lose his design in speaking, what thus spoken, they could not understand." Our knowledge of the chronology of composition of the parts of the two works in question makes it likely that Locke was working on both at the same or roughly the same time. (2) The rejection of the Adamic language was characteristic of Socinian doctrine. Wolfgang Kayser has observed: "So weit ich sehen kann, wird in jener Zeit [he has quoted a work from 1612] allein von der Sekte der Socinianer abgelehnt, von der Tatsache der Namengebung Rückschlüsse auf Adams Weisheit zu ziehen. Auch dass er besonderes Erkenntnisvermögen besessen hätte, wird geleugnet." "Böhmes Natursprachenlehre und ihre Grundlagen," in *Euphorion*, vol. 31 (1930), p. 538. This argument is presented in Fausto Socino's *De Statu Adamantis ante Lapsum*. Its full theological implications are presented in Socino [interpretation of John 1:1-15 in his *Explicationis primae partis primi capituli Johannis*]: "Vers. 1. *In principio erat verbum*. Qui hoc loco principii nomine Christi aeternitatem designare volunt, manifestissimi erroris, vel ex eo solo convincuntur, quod nulla seu Veteris, seu Novi Foedere auctoritate eorum opinio fulciatur. Etenim nusquam reperies in Sacris Literis, principium pro aeternitate usurpari. Qua propter nomen principii in his verbis, non aeternitatem, sed ordinem earum rerum respicere dicemus, quas Johannes de Jesu Christo dilectissimo Dei filio scripturus est, imitatus hac in re Moysen, qui, Historiam suam describens, ab hoc nomine principii & ipse exordium sumpsit *Gen. 1:1* . . . Ita & Johannes dum in *principio* dixit, ad earum rerum initium respexit, de

argument, and that its full meaning can only be recovered with some labor. Posterity lost the key to its understanding, but there can be no doubt that Locke's "half a dozen" friends did possess the key. Our understanding of the *Essay* depends, in turn, on our success in regaining their knowledge, chiefly from Boyle. But we can also indirectly add to it by an examination of Leibniz' responses to some fundamental aspects of Book III.

## VI

Leibniz and Locke agree on several basic points. Book III opens with the well-known words: "God having designed Man for a sociable Creature, made him not only with an inclination, and under a necessity to have fellowship with those of his own kind; but furnished him also with Language, which was to be the great Instrument, and common Tye of Society." Having himself often stated the same conviction, Leibniz assents in typical fashion: "Je suis rejoui de vous voir éloigné du sentiment de M. Hobbes." He also agrees that words have two functions: For the recording of our own thoughts as in memory and for the communication of our thoughts to others. This Leibniz had likewise said before, though in his answer he points to other kinds of useful signs, such as those of algebra, clearly intending to remind Locke of the combinatorial characteristics. They further agree on a practical matter, namely that language must work with "ease and dispatch" or "ease and quickness." But the agreement does not last beyond the first chapter.

At the very beginning of III, ii, Locke lays down the principle that is absolutely fundamental to his discussion of language, which had to open with a firm dismissal of any notion of natural language. He said:

*Words . . . come to be made use of by Men, as the Signs of their Ideas; not by any natural connexion, that there is between particular articulate Sounds and certain Ideas, for then there would be but one Language amongst all Men; but by a voluntary Imposition, whereby such a Word is made arbitrarily the Mark of such an Idea. The use then of Words, is to be sensible Marks of Ideas; and the Ideas they stand for, are their proper and immediate Signification.*

Language, then, is conventional, not natural; words bear no intrinsic relationship to things, but to ideas only—as we are often reminded by Locke's knot-and-bundle metaphor. The meaning of the word has nothing to do with its sound. Locke takes great care to be consistent in this matter and often restates his principle in passages that strike directly at the natural language doctrine. To take just one example, though a very important one, he realized that his argument against our being capable of gaining knowledge of the real essence of substances would collapse if language could in any manner be considered natural. Two passages will illustrate this point. Book III, iv, 11, says:

*Simple Ideas . . . are only to be got by those impressions, Objects themselves make on our Minds, by the proper Inlets appointed to each sort. If they are not received this way, all the Words in the World, made use of to explain, or define any of their Names, will never be able to produce in us the Idea it stands for. For Words being Sounds, can produce in us no other simple Ideas, than of those very Sounds; nor excite any in us, but by that voluntary connexion, which is known to be between them, and those simple Ideas, which common Use has made them Signs of. . . . In Light and Colours, and all other simple Ideas, it is the same thing: for the signification of Sounds, is not natural, but only imposed and arbitrary.*

Similarly, in III, ix, 4, Locke says:

Now since Sounds have no natural connexion with our *Ideas*, but have all their signification from the arbitrary imposition of Men, the *doubtfulness* and *uncertainty of their signification . . .* has its cause more in the *Ideas* they stand for, than in any incapacity there is in one Sound, more than in another, to signify any *Idea*: For in that regard, they are all equally perfect.

Not to have dismissed the last remnant of the doctrine of natural language would have allowed innateness to creep in by the back door. Thus Book III repeats the argument of Book I, and it is reasonable to surmise that Locke felt compelled to devote a separate book to words owing to the increased knowledge of contemporary language doctrines that he gained during the 1670's and especially during the years in France.<sup>64</sup>

quibus sermonem instituerat, non autem omnium, quae multo ante extiterunt." *Opera Omnia*, vol. I (Irenopoli [Amsterdam], post annum Domini 1656), pp. 77–78. This involves the characteristic Socinian rejection of the doctrine of the consubstantiality of the word.

<sup>64</sup> In his *La Philosophie Générale de John Locke* (Paris, 1908), Henry Ollion, without further elaboration, remarks: "Quoi qu'il en soit, Locke a certainement considéré la doctrine de l'innéité comme fondée sur un abus de mots" (p. 71). This, however, may merely refer to the observation that Locke frequently makes, as in III, ii, 5, that "because *Men* would not be thought

Locke's statement of the conventionalist doctrine elicits the longest single passage spoken by Theophile in the entire *Nouveaux Essais*, no less than seven pages punctuated only twice by half-line interruptions from Philalethe—one of which is "Encor un exemple le feroit mieux entendre." The length of the answer was to be expected. Locke's doctrine has brought both to the very center of their disagreement; Leibniz' long and detailed answer therefore offers an especially fruitful opportunity for further exploration and identification of their positions.

Theophile's answer to III, ii, 1 opens:

Je say qu'on a coustume de dire dans les écoles, et par tout ailleurs que les significations des mots sont arbitraires (*ex instituto*) et il est vray qu'elles ne sont point déterminées par une nécessité naturelle, mais elles ne laissent pas de l'estre par des raisons tantost naturelles où le hazard a quelque part, tantost morales où il y entre du choix.

After mention of artificial language, Theophile makes this observation concerning linguistic change: "Pour ce qui est des langues qui se trouvent faites depuis longtemps, il n'y en a gueres qui ne soit extremement alterée aujourd'hui. Cela est manifeste en les comparant avec les anciens livres et monumens qui en restent." He supports his contention with copious citation of examples from his wide knowledge of early documents, many of them legal. In this context he presents an essentially correct division of the Germanic languages and then finally arrives at this important conclusion:

De sorte qu'il n'y a rien en cela qui combatte et qui ne favorise plustost le sentiment de l'origine commune de toutes les Nations, et d'une langue radicale et primitive. Si l'Hebraïque ou l'Arabesque y approche le plus, elle doit estre au moins bien alterée, et il

semble que le Teuton a plus gardé du naturel, et (pour parler le langage de Jaques Böhm) de l'Adamique: car si nous avions la langue primitive dans sa pureté, ou assez conservée pour estre reconnoissable, il faudroit qu'il y parussent les raisons des connexions soit physiques, soit d'une institution arbitraire sage et digne du premier auteur. Mais supposé que nos langues soyent derivatives quant au fonds, elles ont neantmoins quelque chose de primitif en elles mêmes, qui leur est survenu par rapport à des mots radicaux nouveaux, formés depuis chez elles par hazard mais sur des raisons physiques.

He illustrates how this may occur by pointing out that man from the sounds made by frogs has both the Latin "coaxare" and the German "quaken." "Or il semble," he continues, "que le bruit de ces animaux est la racine primordiale d'autres mots de la langue Germanique. Car comme ces animaux font bien du bruit, on l'attribue aujourd'hui aux diseurs de rien et babillards qu'on appelle *quakeler* en diminutif." But these noises also give evidence of life, and "de là est venu que *quak* en vieux Allemand signifioit vie ou vivant, comme on le peut remarquer dans les plus anciens livres, car *quaksilber* est vif-argent, et *erquicken* est conforter et comme revivifier ou recreer après quelque defaillance ou quelque grand travail." He concludes: "Ainsi on peut juger qu'à l'égard de ces mots la langue Germanique peut passer pour primitive; les anciens n'ayant point besoin d'emprunter d'ailleurs un son qui est l'imitation de celui des grenouilles. . . . Car il semble que par un instinct naturel les anciens Germains, Celtes, et autres peuples apparentés avec eux, ont employé la lettre R pour signifier un mouvement violent et un bruit tel que celui de cette lettre."<sup>65</sup> But if "r" suggests violent movement, "l" suggests a gentler one. Thus a child who finds "r" too harsh will substitute "l," saying "mon lèvelend pele" for "mon reverend pere," the characteristic quality

to talk *barely* of their own Imaginations, but of Things as really they are; therefore they often suppose their Words to stand also for the reality of Things." But the passages quoted above can only refer to the abuse that arises from acceptance of the natural language doctrine. It should be remembered that Locke's own title for Book I, as evidenced by the first five editions, is simply "Of innate Notions" and not, as Fraser has it "Neither Principles nor Ideas are innate." "Notion" is no synonym for "idea," and much of Locke's argument, and its right understanding, depends on his introduction and use of the latter by contrast to the former. We have already seen that John Webster in his exposition of the Adamic language uses the terms "internal notions impressed" and "innate notions."

<sup>65</sup> In England, George Dalgarno had made similar observations in his *Didascalocophus* (1680): "All languages guided by the instinct of nature, have more or less of Onomatopoeia in them, and I think our English as much as any. . . . Take for example, wash, dash, splash, flash, clash, hash, lash, slash, trash, gash, &c. So grumble, tumble, crumble, jumble, fumble, stumble, bumble, &c., of which kind of words, the learned, and my worthy friend Dr. Wallis has given a good account in his English Grammar. In all these and such like words there is something symbolizing, and analogous to the notions of the things, which makes them both more emphatic and easy to the memory. But in words literally written, and of a meer arbitrary institution, there can be no thing symbolical." *Works of George Dalgarno* (Edinburgh: Maitland Club Reprint, 1839), p. 126. Leibniz was familiar with both Dalgarno's work and with John Wallis' *Grammatica Linguae Anglicanae*. Schottel's *Ausführliche Arbeit* presents a wealth of similar observations and examples.

of "I" also being illustrated in "leben," "lind," "lentus," "lieben," "lauffen," etc. Again, this proves "qu'il y a quelque chose de naturel dans l'origine des mots qui marque un rapport entre les choses et les sons et mouvemens des organes de la voix."<sup>66</sup> Leibniz is here using the affective theory of the sound-thing connection, which has always played a major role in attempts to explain the origin of language. It was in fact in this manner that he believed Adam had found names for the animals: "At in linguis paulatim natis orta sunt vocabula per occasiones ex analogia vocis cum affectu, qui rei sensum comitabatur: nec aliter Adamum nomina imposuisse crediderim."<sup>67</sup>

In this fashion Leibniz sought to rescue language—and the consistency of his own philosophy—from Locke's doctrine that words "signify only Mens peculiar Ideas, and that by a perfectly arbitrary Imposition" (III, ii, 8). Some of Leibniz' observations and many of his examples are not his own, Schottel's *Ausführliche Arbeit*: being his most frequent source. But the total integration into his own philosophy is of course entirely original with him; and hence the very long answer to Locke's initial statement in III, ii is not a digression, but essential to the maintenance of Leibniz' over-all philosophy. In regard to the nature of language,

however, Locke is certainly more radical and original than Leibniz; and there can, as we have seen, be no doubt whatever that Locke very well knew he was arguing against the sort of position that Leibniz later adopted in his Critique. In general, Locke's saying that he did not feel he sufficiently understood Leibniz' early pieces on the *Essay*, may have been a polite—or circuitous—way of saying that he was surprised that Leibniz had understood so little.<sup>68</sup>

Toward the end of the answer to III, ii, I, Leibniz, understandably makes some remarks about etymology, a subject that Locke does not appear to have taken any interest in just as he never to my knowledge uses the word itself.

Les langues en general estant les plus anciens momumens des peuples, avant l'écriture et les arts, en marquent le mieux l'origine, cognations et migrations. C'est pourquoy les Etymologies bien entendues seroient curieuses et de consequence, mais il faut joindre des langues de plusieurs peuples, et ne point faire trop de sauts d'une nation à une autre fort éloignée, sans en avoir des bonnes verifications, où il sert sur tout d'avoir les peuples entre deux pour garans. Et en general l'on ne doit donner quelque creance aux etymologies que lors qu'il y a quantité d'indices concourans.<sup>69</sup>

<sup>66</sup> This statement should be closely compared with John Webster's statement quoted above at footnote 56. Leibniz' examples are designed to demonstrate that the principles Webster attributes to the Adamic language can currently be seen to operate. Cf. this passage from a letter from Leibniz to Sparwenfeld written between April and August 1699: "Il me semble que *n, l, r, s, t*, ont beaucoup de connexion dans leur formation. Ceux qui ne sçauroient bien prononcer *r*, le changent en *l*, disant *mon leoslend père*, au lieu de *mon rooslend père*, comme les florentins changent le *k* en *h*." See Harald Wieselgren, "Leibniz' brief till Sparfvenfelt," p. 53.

<sup>67</sup> *Miscellanea Berolinensia* (Berlin, 1710), p. 2. The "Brevis designatio" was reprinted in L. Dutens, ed., *Leibnitii . . . Opera Omnia* (Genevae, 1768), vol. IV, pt. ii, pp. 186–198.

<sup>68</sup> The reader may have noticed that Philaethe and Theophile often seem to argue at cross-purposes, the inevitable result of Theophile's occasional failure to understand Philaethe's position, or of disagreement so complete that no common ground is left. A notable case involving language occurs in IV, v, of the *Nouveaux Essais*, i.e. after Book III at a point where one might not have expected any misunderstanding. Philaethe has rendered Locke's statement in IV, v, 2 that "Truth properly belongs only to Propositions: whereof there are two Sorts, viz. Mental and Verbal; as there are two sorts of Signs commonly made Use of, viz. Ideas and Words." Theophile answers: "Mais ce que je trouve le moins à mon gré dans votre definition de la Verité, c'est qu'on y cherche la verité dans les mots. Ainsi le même sens, étant exprimé en Latin, Allemand, Anglois, François, ce ne sera la même verité. Et il faudra dire avec M. Hobbes, que la verité depend du bon plaisir des hommes. Ce qui est parler d'une maniere bien étrange." Leibniz is certainly not the only one to misunderstand Locke on this point, and Locke's lack of clarity is in part to blame. But Locke could at least in this case not have meant anything of the sort that Leibniz suggests, for what Locke does mean, though it is not clear from this passage alone, is of course not words as sounds, but words as knots that tie bundles of ideas together. In different languages the same number of ideas may be differently sorted into bundles, and some that are "bundled" in one may not be in another—hence the doctrine of intranslatability among languages. Locke's use of the word "language" rather than "the English language" or "English" has had very considerable consequences. I would very much doubt that Locke always meant to imply that he was talking about language in general every time he used the word "language"; but his usage made it possible for the eighteenth century to read parts of the *Essay* as a discussion of language at large in the manner of the Port-Royal *Logique* and the *Grammaire Générale et Raisonnée*. If Locke had been forced always to make the distinction between "langue" and "langage" as in French, the matter might have been different. Coste, in his translation, normally uses "langage" for Locke's "language." The crossing of these two streams became immensely productive, and the eager eighteenth-century discussion of language, with wide use of Locke, cannot be imagined without it.

<sup>69</sup> Cf. a very similar passage, which Leibniz may well have known, in le P. Besnier, *La Raison des Langues* (Paris, 1674), p. 21: "Puis donc que la corruption des Langues ne s'est faite qu'insensiblement, & peu à peu; il ne faut, pour la découvrir sûrement, que remonter par degrez, jusques à la première source de leurs différences: prenant bien garde à ne point faire de

This warning against careless procedures in etymology is followed by this reference: "Je me souviens que feu Mons. Claubergius philosophe excellent a donné un petit Essai sur les origines de la langue Germanique, qui fait regretter la perte de ce qu'il avoit promis sur ce sujet." Clauberg was a well-known Dutch Cartesian, and the work to which Leibniz refers is his *Ars Etymologica Teutonum*, which he had first become acquainted with in the early months of 1699, though it had appeared in 1663.<sup>70</sup> In this work Clauberg had observed that words for the operations of the intellect originally had reference to sensible things and actions, noting for instance:

... à Lat. *cipio* multa ejusdem significationis vocabula descendunt, uti *percipere*, *concupere*, *conceptus*, "sinnbegriff". . . . Equidem Germ. *nehmen* in tali significatione frequentissimum est: e.g. *abnehmen*, in *acht nehmen*, *wahrnehmen*, *rath nehmen*, *sich etwas fürnehmen*, *er lässt sich vernehmen*.<sup>71</sup>

On this basis he establishes one of his etymological rules: "*A sensibilibus ad intelligibilia quamplurima vocabula sunt traducta*"; for this principle he refers the reader to his *Logica Vetus et Nova* with these words: "Proprie quidem *capere* & *prehendere* sunt manuum actiones . . . at quomodo ad intellectum ejusmodi verba sint ab hominibus traducta, uberius explicui Log. proleg. § 77."<sup>72</sup> This paragraph in the "Prolegomena" to the *Logica* reads:

Quotiescunque igitur bene aliquid à mente intelligi, vel intelligi posse, significare volumus, iis utimur loquendi modis, quibus alioquin proprie organorum corporeorum sensus (quos scil. iis tribuimus) aut

sensibilia eorum objecta designantur: ex: gr à *visu* & rebus visibilibus ad intellectum & res intellectuales illa transferimus: evidens est, evidenter apparet, clarè hoc video, elucet, perspicuum est, perspicax ingenium, vel caeco appareat, sole meridiano clarius, oratio dilucida &c. à *tactu*: teneo, apprehendo, comprehendo, arripio, rem acu tango, attingo, capio, manifestum est &c.

This passage bears a striking resemblance to the well-known opening words of III, i, 5, which states the same principle and uses some of the same examples in the same order:

It may also lead us a little towards the Original of all our Notions and Knowledge, if we remark, how great a Dependence our Words have on common sensible Ideas; and how those, which are made use of to stand for Actions and Notions quite removed from sense, have their rise from thence, and from obvious sensible Ideas are transferred to more abstruse Significations . . . v.g. to *Imagine*, *Apprehend*, *Comprehend*, *Adhere*, *Conceive*, *Instill*, *Disgust*, *Disturbance*, *Tranquillity*, &c. are all Words taken from the Operations of sensible Things, and applied to certain Modes of Thinking.

Locke may not have seen the *Ars Etymologica*, but the probability is very high that he knew the contents of the *Logica*, for it is mentioned along with four other works by Clauberg in the five-page entry called "Methode pour bien etudier la doctrine de M<sup>r</sup> de Cartes," which Locke entered in his Journal on 7 March 1678 while he was in Paris.<sup>73</sup> We further know that Locke bought a copy of the *Logica* round the same time, for on 30 June 1678, it figures in a list of "Books of mine put in the box marked C. B. No. 3." If, indeed, Locke's

fausse démarche, qui puisse conduire un peu trop loin, & engager dans un mauvais pas, d'où l'on auroit de la peine à se retirer. C'est là l'unique moyen que j'aye trouvé, de répandre un certain air de vraisemblance sur tout ce qui regarde cette matière, qui n'a de probabilité qu'autant qu'on luy en donne, en la ménageant si bien, qu'on n'aille jamais d'une extrémité à l'autre, sans passer par le milieu, qui a esté comme le lien de ces deux extrémités. Car c'est de cet enchaînement de mots, & de cette suite d'altérations, que depend principalement toute la justesse & toute la vraisemblance de la méthode." Leibniz had already heard of Besnier's *Reunion* in Paris shortly after it appeared. See letter to Sparwenfeld in Wieselgren, pp. 16-18, dated 29 Jan. 1697. The letter shows that Leibniz was also familiar with Besnier's "Discours sur la Science des Etymologies." In this letter, Leibniz takes a rather critical view of Besnier's work, an opinion shared by Sparwenfeld, who had known Besnier in Paris. The historical uses of etymology had been pointed out by Camden as quoted in Schottel's *Ausführliche Arbeit*, p. 127: "In lingue certissimum originis gentium argumentum: Qui enim lingue societate conjuncte sunt, originis etiam communione fuisse conjunctos, homo opinor nemo inficiabitur. Quod si omnes omnium historie intercidissent, ipsarum linguarum communitas id facile evinceret, imò facilius, quam vel gravissimorum historicorum autoritas."

<sup>70</sup> See Pietsch, p. 322.

<sup>71</sup> See Eccard's *Collectanea Etymologica*, pt. I, pp. 193-194. On Clauberg's work, see N. Scheid, "Eine Philosophische Wortdeutungslehre aus dem Jahre 1663," in *Zeitschrift des Allgemeinen Deutschen Sprachvereins*, vol. 23 (1908), pp. 5-9.

<sup>72</sup> Eccard, pp. 194-195. § 77 in the "Prolegomena" to Clauberg's *Logica* occurs only in the second and subsequent editions: Jon. Claubergii *Logica Vetus & Nova*, modum inveniendae ac tradendae veritatis, in genesi simul & analysi, facili methodo exhibens. Editio secunda mille locis emendata novisque Prolegomenis aucta. Amstelædami, ex Officina Elzeviriana, 1658. In this edition, the "Prolegomena" occupy pp. 1-37 with a total of 103 paragraphs divided in five chapters. The first edition appeared in 1654, its "Prolegomena" had only 37 paragraphs, and total number of pages was 413 against 463 for the second edition.

<sup>73</sup> Aaron & Gibb, pp. 105-111.

source is the *Logica*, then we gain valuable information not only about the composition of the *Essay* and its chronology but also about his close knowledge of a particular work that commented on matters of language. We are, furthermore, again encouraged to assume that the years in France exposed Locke to much new material, which in turn influenced the *Essay* and helped determine the form it ultimately came to assume. The *Essay* itself gives few clues to Locke's reading and specific knowledge, but our right reading of it does in large measure depend on our success in recovering as much of that knowledge as we can.<sup>74</sup>

Leibniz gives a characteristic answer to the opening words of III, i, 5, to Locke's assurance, as stated later in the same section, that we shall "find, in all Languages, the Names, which stand for Things that fall not under our Senses to have had their first rise from sensible *Ideas*." Theophile says:

C'est que nos besoins nous ont obligé de quitter l'ordre naturel des idées, car cet ordre seroit commun aux anges et aux hommes et à toutes les intelligences en general, et devroit estre suivi de nous, si nous n'avions point égard à nos interets: il a donc fallu s'attacher à celui que les occasions et les accidens où nostre espece est sujette nous ont fourni, et cet ordre ne donne pas l'origine des notions, mais pour ainsi dire l'histoire de nos découvertes.

The first part of this passage makes implied reference to man's banishment to the world of sense after the Fall. The concluding argument depends on the doctrine of harmony. The "origine des notions" implies, as Locke no doubt intended in his corresponding words, that the understanding alone has made those notions by the pattern "Nature . . . unawares suggested to Men," whereas "l'histoire de nos découvertes" suggests that it is not a matter of making or inventing according to some more or less chance pattern, but of discovering what is already there, contained within the order of the pre-established harmony. But Locke would have been indifferent to Leibniz' distinction, and his practical—one might almost say pragmatic—

insistence on objectivity forced him to dismiss the doctrine that some over-all harmony enveloped man and nature. Locke's method was the "historical, plain method," not the speculative method that could afford to operate with concepts of the possible and the contingent. Two of Locke's examples were that "angel" originally meant "messenger" and "spirit" meant "breath." Leibniz sensed dangerous impiety in those examples and implied a warning by telling of a certain Dutchman who had abused "cette verité (que les termes de Theologie, de Morale et de Metaphysique sont pris originairement des choses grossieres) pour tourner en ridicule la Theologie et la foy chrestienne dans un petit dictionnaire Flamand où il donnoit aux termes des definitions ou explications, non pas telles que l'usage demande, mais telles que sembloit porter la force originaire des mots, et les tournoit malignement." This impiety had cost its author a term in the workhouse. But with Locke the argument does not stick, for he was merely illustrating that we grasp the less familiar by the more familiar, the "Actions and Notions quite removed from sense" by those that are not. This view is not impious to men who shared the Royal Society's conviction that Nature was the great "unsealed book of God." In England piety could be exercised in a manner that on the Continent was considered downright materialism. Leibniz was shocked and disturbed, but Condillac and the French *philosophes*, who drew the same inference, were delighted half a century later. It is perhaps not untrue to say that one of the great reversals in modern European thought occurred by misunderstanding rather than by the more normal avenue of the consistent pursuit of ideas and their logical consequences.

In historical terms, we are left with the paradox that Locke's *Essay* throughout the eighteenth century exercised an influence on language study which I am sure was as foreign to his intentions as it would, with certain important qualifications, have been proper to those of Leibniz, who often voiced the opinion we find at the end of III, vii, 6:

<sup>74</sup> Peter Laslett has since kindly informed me that Clauberg's *Logica* appears on two occasions after Locke had purchased it in 1678; it was not in Locke's final library. It is item 736b in the forthcoming catalogue of Locke's library that is being prepared by Messrs. Laslett and J. R. Harrison. In his *Reunion des Langues*, Bernier makes an observation that is similar to Locke's and Clauberg's. He notes that "la plupart de nos paroles si on les compare à leur première origine, ne sont que des metaphores." In this context he remarks: "Ainsi, quand nous disons que l'esprit, ou l'entendement s'applique à penser, à concevoir, à discourir, à expliquer, à débrouiller, à démealer les matières, à découvrir la vérité; quand nous parlons du trouble, des aversions, de l'agitation, & de la consternation de la volonté; pour expliquer les actions du monde les plus spirituelles, nous nous servons d'images, qui sont en effet corporelles dans leur première origine, quoy-qu'elles ayent la plupart perdu leur signification propre, pour en prendre une autre purement figurée" (pp. 38-39).

Je croye veritablement, que les langues sont le meilleur miroir de l'esprit humain, et qu'une analyse exacte de la signification des mots feroit mieux connoitre que toute autre chose, les operations de l'entendement.

This bears much more resemblance than anything in Locke, to Turgot's conviction that etymology is "une branche intéressante de la métaphysique expérimentale" or to his optimistic battle cry that "ceux qui observent la marche de l'esprit humain . . . doivent marcher sans cesse le flambeau de

l'Etymologie à la main," in spite of Turgot's own claim "je parle d'après Locke."<sup>75</sup> But Leibniz' *Nouveaux Essais* lay unpublished in the library at Hanover until 1765, and by that time the *Essay* had been made to yield very nearly the same doctrines about etymology and the origin of language. To Locke language was only a tool, its discussion only incidental to his examination of our ways of lapsing into error and becoming diverted from certain knowledge. But to posterity words and language became, contrary to Locke's aims, Leibniz' mirror of the mind.<sup>76</sup>

Princeton University

<sup>75</sup> What is true of Turgot and the *philosophes* is also true of James Mill and the Utilitarian psychology he presents in his *Analysis*. See my article on "The Early History of the *Oxford English Dictionary*," in the *Bulletin of the New York Public Library*, vol. 66 (September, 1962), pp. 417-439.

<sup>76</sup> It is a misunderstanding of the same sort to believe that Locke in the *Essay*—chiefly in Book III, but also elsewhere—proposed to give an exhaustive and systematic analysis of language. He was not dealing with language as a linguist or an etymologist, but as a philosopher who had certain practical concerns in mind, and they did not embrace all of language, and least of all in the manner we may nowadays conceive of such an enterprise. In his comment on III, ii, 1, R. I. Aaron says in his *John Locke*, pp. 208-209: "Unfortunately, Locke does not here examine those words which are obviously onomatopoeic, nor again words derived from an onomatopoeic source (as Leibniz did in discussing this passage). Indeed, he makes no attempt to examine the historical origins of language, though this would have been in accordance with his own expressed method of procedure. It is true that etymology was in its infancy at the time, but Locke does not make use even of the limited information then available." I am sure that Locke was not at all interested in examining the "historical origins of language," and least of all as we understand that phrase. His unacknowledged disciples were, but Locke only used seemingly etymological observation in a simple, commonsense and practical fashion to support a point or make it plausible, but he never made an argument depend on it; he was interested in the origin of our notions, but not in the origin of language. Onomatopoeia might have amused him, but I think he would also have found it rather trivial, a quirk rather than the source of a principle; after all, the question of the real or nominal essence of gold is not illuminated by onomatopoeic considerations. We have been told again and again that etymology at some point in time was in its infancy—it is the constant nineteenth-century phrase for the state of etymology in the eighteenth century (which is a gross injustice)—but I have never heard whether it has since been in its puberty or adolescence or is now in its senility. "Etymology" is one of the trickiest words of all, both philosophically and philologically, and nineteenth-century pride in its own accomplishments in that department should not deceive us.

## II. A NEW LOOK AT HERACLITUS

CHARLES H. KAHN

NOVELTY is a relative concept, and in a sense the Preacher is surely right: there is nothing new under the sun. This seems particularly true in the case of an author like Heraclitus, who has been a subject for the very frequent making of books (and articles) for well over a century. Nor does the flood of commentary show any signs of abating.<sup>1</sup> In proposing, therefore, to take a new look at Heraclitus, I certainly do not pretend to offer a picture entirely different from everything that anyone else has ever seen. It is possible that one or two of my predecessors—and notably Reinhardt—might have recognized my “new” view as only a logical development of their own. But the interpretation to be put forth here does, I think, contrast with much recent work on Heraclitus, particularly in English,<sup>2</sup> and, in one respect at least, it diverges from the mainstream of Heraclitean scholarship since Diels. This point of disagreement, with which I shall begin, concerns the literary form and hence the semantic status of Heraclitus’ discourse: if my view is correct, many commentators are in danger of radically mistaking the nature of his work and the quality of his thought. From this preliminary question of form I shall proceed to the philosophic content of the discourse, passing in review its major themes. The aim throughout will be to present, not a detailed exegesis of individual fragments, but an over-all framework of interpretation into which every fragment will fit. The presupposition of this approach is that piece-meal discussion of individual texts may become arbitrary and pointless unless it is based upon a sound appreciation of Heraclitus’ distinctive character as thinker and as stylist. The interpretation will

conclude with the topic I take to lie at the center of all Heraclitus’ thought and concern: the nature of the human soul and its relationship to the divine principle of unity in the cosmos.

### I. THE BOOK

What kind of a book did Heraclitus compose? Or was it a book at all? Is it possible to form any general idea of a work from which we possess some 130 random quotations?

Early editors, such as Bywater, tried to group the fragments by subject matter.<sup>3</sup> Since 1901, however, the standard arrangement has been that of Diels, which lists the fragments alphabetically, according to the name of the author who happens to cite them. This apparently irrational procedure may be justified on sound philological grounds. Recognizing that any arrangement by subject matter was to some extent arbitrary, Diels wished above all to avoid imposing any personal interpretation upon his edition of the texts. In fact, however, by the atomistic character of his arrangement, he has largely succeeded in imposing his own view of Heraclitus’ work as lacking in literary structure.<sup>4</sup> For Diels was motivated not only by the difficulty (or impossibility) of reconstructing the original sequence of the fragments. He also called attention to their aphoristic style, their resemblance to the gnomes or sayings of the Seven Sages, and (with Nietzsche’s *Zarathustra* in mind) he suggested that these sentences had originally been set down in a kind of notebook or philosophical journal, with no literary form or unity linking them to one another. He thus implied, after

<sup>1</sup> For a recent survey of the secondary literature see the vastly expanded translation of Zeller’s work by R. Mondolfo, *La filosofia dei Greci nel suo sviluppo storico*, pt. I, vol. IV (Florence, 1961), pp. 1–6. For the impact of Heraclitus on modern authors from Hegel to T. S. Eliot see K. Axelos, *Héraclite et la Philosophie* (Paris, 1962). There is a new and important study of selected fragments in J. Kerschenshteiner, *Kosmos, Zetemata* 90 (Munich, 1962), pp. 97–114.

<sup>2</sup> See in particular G. S. Kirk, *Heraclitus: the Cosmic Fragments* (Cambridge, 1954), and his summary statement in Kirk and Raven, *The Presocratic Philosophers* (Cambridge, 1957). Kirk’s views have had a considerable influence on W. K. C. Guthrie’s treatment of Heraclitus in *A History of Greek Philosophy*, vol. I (Cambridge, 1962).

<sup>3</sup> I. Bywater, *Heracliti Ephesii reliquiae* (Oxford, 1877). In many respects this has remained the most useful edition for any detailed study of Heraclitus.

<sup>4</sup> See, for example, K. Axelos, *op. cit.*, p. 248: “la pensée d’Héraclite, globale, mais aphoristique et fragmentaire par essence” (italics added).

all, that the chaotic pattern of his arrangement gave a true picture of Heraclitus' own composition. In the case of Heraclitus, arrangement and interpretation are in fact inseparable from one another,<sup>5</sup> as Diels saw in the work of his predecessors. His mistake was to imagine that his own order could be an exception.

Our interpretation will be based upon the opposite assumption; that Heraclitus' discourse as a whole was as carefully and artistically composed as are the preserved parts, and that the formal ordering of the whole was probably as much an element in its total meaning as in the case of any lyric poem from the same period. The true parallel for an understanding of Heraclitus' style is, I suggest, not Nietzsche but his own contemporaries, Pindar and Aeschylus. The extant fragments reveal a command of word-order, imagery, and studied ambiguity as effective and as self-assured as that to be found in any work of these two poets. We can, I think, best imagine the structure of Heraclitus' little book on the analogy of the great choral odes, with their fluid but carefully articulated movement from image to aphorism, from myth to riddle to contemporary allusion. Yet the intellectual unity of Heraclitus' work was certainly greater than that of any archaic poem, since its final intent was more explicitly didactic; in fact, its central theme was precisely an assertion of the principle of unity: *hen panta einai*, "all things are one." The content of this very general formula seems to have been filled in by a coherent chain of statements linked together not by logical argument but by interlocking ideas and verbal echoes, with an elaborate use of imagery, word-play, and enigma. Theophrastus found the result "incomplete and contradictory,"<sup>6</sup> but he was looking for a prosaic exposition of physical theories. Heraclitus is not merely a philosopher but a poet as well, and, furthermore, one who

chose to speak in tones of prophecy. The literary effect he aimed at was scarcely that of didactic prose; it may better be compared to the impact of Aeschylus' *Oresteia*: the solemn and dramatic unfolding of a great truth, step by step, where the sense of what has gone before is continually modified and enriched by its echo in what follows.

That Heraclitus' discourse possessed an artistic design of this type can scarcely be demonstrated, but it is made *a priori* probable by the clear evidence of literary artistry in every fragment where the original wording has been preserved intact. Furthermore, indirect evidence for a larger structure is provided by the impossibility of interpreting many of his statements in isolation. For example, how are we to understand fragment 60: "The way up and down is one and the same"? The literal interpretation of this remark poses no difficulties, but taken in isolation it is so ambiguous as to be devoid of any significance. Everything turns upon what kind of "way" is meant, and that we could only learn from the context—from the sentences that came before and after. And the case of this fragment is typical: unless we provide them with a context, many of Heraclitus' statements cannot be interpreted in any significant way. This suggests one of two things: either these ambiguous fragments originally had no definite sense, or Heraclitus himself gave them a significant context.

There is of course no chance of restoring these shattered and incomplete fragments to the order in which Heraclitus himself disposed them. But once we assume that Heraclitus' ordering had some plan, it becomes the interpreter's task to arrange them in the most meaningful pattern he can find.<sup>7</sup> Not only does arrangement imply interpretation; interpretation, in turn, involves arrangement; and the methodical complement

<sup>5</sup> Thus Kirk, who holds to Diels's view of the composition of the fragments, nevertheless groups them by subject matter in order to discuss their meaning, and rightly remarks that his "own interpretation of a fragment is apparent from the very beginning, from the group in which it is placed" (*Heraclitus: the Cosmic Fragments*, p. xii). Kirk goes beyond Diels in supposing that there was originally no book at all, but a set of famous "sayings" or oral utterances, collected, perhaps, by a pupil (*op. cit.*, p. 7). Aside from the need for assuming the unattested (and, on the whole, unlikely) existence of a pious disciple, Kirk's hypothesis is contradicted by the only relevant bit of external evidence, the story that Heraclitus himself deposited the book in the temple of Artemis (Diogenes Laertius, IX. 9).

<sup>6</sup> Quoted by Diogenes Laertius, IX. 6.

<sup>7</sup> Burnet (who wisely preferred Bywater's sequence) was one of the few classical commentators to protest against the principle of Diels's arrangement. An even more striking dissent from Diels's view is the unjustly neglected article of H. Gomperz, "Über die ursprüngliche Reihenfolge einiger Bruchstücke Heraklits," *Hermes*, vol. 58 (1923), p. 20. Although my own conclusions were formulated before encountering Gomperz's study, my opposition to Diels is based upon much the same principles as his. A comparable view is that of Hermann Fränkel, who emphasizes the care and artistry which is characteristic of the preserved fragments and remarks, "Der Bau des ganzen Werkes wird entsprechend gewesen sein" (*Wege und Formen Frühgriechischen Denkens*, p. 78, n. 3). Fränkel thinks, however, that the task of reconstruction is hopeless (*Dichtung und Philosophie*, p. 474).

to a new interpretation of Heraclitus would be a complete reordering of the fragments to replace the meaningless sequence of Diels. I hope to present such an arrangement on another occasion; here we must be content to state the principle of contextual juxtaposition as an element in the total meaning of Heraclitus' words. There is more to be said about the literary quality of the work—in particular, concerning Heraclitus' use of paradox, riddle, and verbal ambiguity. But these stylistic features may conveniently be discussed in the context of his thought as a whole.

## II. THE "LOGOS": STYLE AND SUBSTANCE

The thought of Heraclitus constitutes a fully articulated vision of the world, one which is ultimately to be understood in its own terms. But in order to penetrate within the Heraclitean universe, we must first locate it against a definite historical background. This background may be defined by reference to three or four fixed points. In the first place, Heraclitus takes for granted the Milesian view of the natural world as a cosmos, an organized structure of elemental principles and opposing powers, functioning in accordance with a cyclical order of time. In the second place, he is familiar with the new conception of deity as mind or *nous*, formulated above all by Xenophanes, a conception developed in close connection with natural philosophy and in conscious opposition to the Homeric portrayal of the gods.<sup>8</sup> Furthermore, he is familiar with another decisive innovation, the new view of the human soul as potentially immortal and hence divine, a view associated above all with the name of Pythagoras. Finally, Heraclitus adopted from his predecessors the notion that the universe is pervaded by the principle of mathematical measure or proportion, and that this principle is strikingly embodied in the musical attunement or *harmonia* of the strings of a lyre. A similar mathematical conception is implicit in the Milesian conception of the cosmos, but the form which it takes for Heraclitus is probably influenced by Pythagorean speculation on the relationship between musical intervals and numerical proportion.

This triple influence upon Heraclitus is reflected in the fragments by his personal attacks on Xenophanes and Pythagoras, and on Miletus in the

person of Hecataeus (frs. 40, 81, 129); other thinkers have been known to attack men to whom they are indebted. But to say that Heraclitus was influenced by his predecessors is not to say that he agreed with them. More interesting than the fact of influence itself is the nature of the transformation which earlier ideas have undergone as a result of their appropriation by Heraclitus. The points of contact have too often been interpreted as indications that Heraclitus and his predecessors are concerned with the same questions, that their thought is as it were on the same level. On the contrary, he clearly expresses his contempt for their *polymathia*, the mere erudition which does not lead to significant insight. In Heraclitus' view, men like Xenophanes and Pythagoras failed to see the true meaning of their own knowledge. Whatever he has taken over from them must mean something very different for him.

The distinctive character of Heraclitus' own view is suggested, at the very outset of his discourse, by an emphatic reference to the term for discourse itself:

Although this word (*logos*) is true forever, men ever fail to comprehend, both before hearing it and when they hear it first. Though all things come to pass according to this *logos*, men are like the untied when they try such words and deeds as I set forth, sorting out each thing according to its nature and declaring how it is. (fr. 1)

The term *logos* comes to us laden with the whole history of Western philosophy and theology. To avoid anachronism, and to seize the meaning of *logos* in all the concreteness and complexity which it possessed for Heraclitus, we must briefly recall the early history of the term before it came to mean the rational faculty of the soul, the Reason of the Universe, or the Word of God.

The word *logos* is a verbal noun, whose primitive meaning is inseparable from that of the verb *legein* to which it corresponds. In its most current, classical sense *legein* means "to say" and *logos*, accordingly, means "saying, utterance, speech"; as in the standard antithesis of *logos* and *ergon*, "word" and "deed." But *logos* always involves more than mere speech, and this added element of meaning largely stems from the original force of the verb. In its earliest usage (for example, in Homer) *legein* means either (1) to pick up or

<sup>8</sup> For the points of contact between Heraclitus and Xenophanes, see above all O. Gigon, *Untersuchungen zu Heraklit* (Leipzig, 1935).

gather, or (2) to count or enumerate.<sup>9</sup> The underlying idea, reflected in both senses, seems to be "to group or gather, passing from one thing to another."<sup>10</sup> To the early use of the verb for counting corresponds the frequent and classical sense of *logos* as "computation, reckoning, account." The notion of enumeration also gives rise to the more specifically mathematical meaning of *logos*, as "ratio" or "proportion": a geometric ratio may be seen as a special case of the grouping of things for counting or computation.

On the other hand, the verb *legein* "to count," like the English "to tell" or the French *compter*, may pass over into the sense of "to recount, narrate," *conter*, and hence (again like "tell") into the more general sense of "to say." In this way, the noun *logos* came to mean a "tale" or "narration," and, by extension, any act of speaking or saying.<sup>11</sup> But as *legein* means not merely to talk, but to say something significant,<sup>12</sup> so *logos*, even when it denotes "speech," always refers to the meaning and content of the words as well. Because of its other uses, it may also carry with it some connotations of rational collecting, arranging, or enumerating; and it is these collateral meanings which explain the extraordinary importance of the term in Greek philosophy. The early poets may use *logos* and *mythos* as synonyms, to denote any speech or story. But when the philosophers undertook to distinguish between a rational account and an old wives' tale, they naturally opposed *logos*, as reasoned, meaningful discourse,

to *mythos*, a merely legendary tale. And so it was that the distinction between "myth" and "reason" was born.

Returning to Heraclitus, we can see that his *logos* is first and foremost his concrete utterance, but that his discourse is never a matter of mere words. The *logos* has a content, a sense and structure that most men do not grasp even when their ears are listening to the words: that is what it means to say that they have "barbarian souls," which do not understand the language of the *logos* (fr. 107). And the content of Heraclitus' discourse is neither more nor less than the world-order which he asserts: "All things come to pass according to this *logos*" (fr. 1).

Thus, from the very first sentence of his book, Heraclitus' use of the word *logos* reflects the fundamental ambiguity of the term: on the one hand, a specific utterance, on the other hand, an orderly relationship between things which is reflected in discourse, including the quantitative relationship reflected in a calculation or ratio. The quantitative sense is unmistakable in several uses of *logos* in the extant fragments: sea is "measured into: the same *logos* as before it became earth" (fr. 31); the *logos* of the soul is so "deep" that we cannot find its end-terms or limits (fr. 45); soul has a ratio or account which is self-augmenting (fr. 115). In virtue of this ambiguity, systematically maintained, Heraclitus' use of the term *logos* for his "discourse" implies at the same time an allusion to the quantitative "measures" that mark the world-order, the

<sup>9</sup> For a full discussion, see H. Boeder, "Der frühgriechische Wortgebrauch von Logos und Aletheia," *Archiv für Begriffsgeschichte*, vol. 4 (1959), p. 82. However, the etymological background is confused rather than clarified by Boeder's linking of the two distinct roots \**leg-* (*λέγω*) and \**legh-* (*λέχομαι*). For a survey of earlier discussions of the term *logos* see Kirk, *Heraclitus*, pp. 37-40. My own view is perhaps closest to that of E. L. Minar, "The Logos of Heraclitus," *Classical Philology*, vol. 34 (1939), pp. 323 ff. But no one, as far as I can see, has noted the essential importance for Heraclitus of significant word-play, as a means of saying several things at once. Hence much fruitless debate about the "true" or primary meaning of *logos* in the fragments. In fact Heraclitus uses the word with different senses in different contexts, but each use is designed to allude to the others at the same time. Compare the two non-synonymous examples of *logos* in fr. 1.

<sup>10</sup> This original sense of the verb is not only Greek but Indo-European, and is reflected in the Latin cognate, *lego*, *legere*, "to gather," "to choose," and hence "to read." The explanation of the latter usage is apparently that, for an ancient, "to read" means to gather meaningful groups (i.e., words) as one passes along the unspaced line in which letters were written one after another.

<sup>11</sup> It is a curious literary accident that the use of *logos* in the early epic is restricted to this general meaning of "utterance," whereas the more etymological sense of "number, account" is current in classical prose and poetry. For a more complete listing of the early uses of *logos*, see Guthrie, *History*, vol. I, pp. 420-24. All of Guthrie's examples can, I think, be explained on the basis of the three primary meanings: (1) gathering, collection, (2) account, enumeration, (3) narration, statement in words. Thus, the idea of "esteem, reputation" is derived by extension from sense (2): to esteem is to take account of, to rate (highly) in one's calculation of a man's worth. The notion of "cause, explanation" (as in Leucippus, "everything occurs from *logos* and by necessity") is also closely related to that of rational enumeration: to explain something is to give an account of it (*logon didonai*). Since the "account" is normally rendered in words, there is an inevitable blending of sense (2) and (3). In the Herodotean examples of "common *logos*" (e.g., I. 141. 4; 3.119.1) the etymological unity of the three senses seems almost to have been re-established: a common *logos* is a gathering of men who are counted together on the basis of a *homologia*, a mutual understanding in words.

<sup>12</sup> As in the idioms *λέγειν τι*, "to make a point"; *οὐδὲν λέγειν*, "what you say makes no sense."

measures which the Sun must respect (fr. 94), and according to which the cosmic Fire is lighted and quenched (fr. 30). Thus, word-play for Heraclitus becomes not so much a literary mannerism as a revelation in language of the hidden unity of the universe, a hint of the orderly structure which his *logos* evokes.

Now the term *logos* is itself peculiarly suited to this type of meaningful word-play and equally apt as an expression of the profound unity of things. For as we have noted, the etymological sense is that of a gathering-together, and above all, of a rational collecting of things by number and sequence. In the linguistic usage of Heraclitus' day, the connotations of gathering had become dormant, but they could easily be revived in certain compounds and combinations, as for example in the expression *koinos logos*, applied to men who make common cause and are to be counted together.<sup>13</sup> Hence when Heraclitus says that "although the *logos* is common, most men live as if they had a wisdom of their own" (fr. 2), he is playing on the sense of gathering-together which is latent in the term *logos*; just as, when he says "it is wise, listening not to me but to the *logos*, to agree that all things are one" (fr. 50), he is playing on the sense of "common *logos*" by his use of the term *homologeîn*, "to agree": literally "to come together in *logos*, to say the same thing." Thus, behind Heraclitus' concrete use of *logos* for his own discourse, lie two interrelated senses for the gathering-together of things: on the one hand, *logos* as human speech points to the coming together of many meanings in words and word-play; on the other hand, *logos* as an "account" of the world points to the convergent unity of things, computed by their measure and proportion. Puns and ratios are equally legitimate devices for revealing the latent *harmonia*, the hidden "fitting-together" of things in word and fact (fr. 54). Heraclitus' emphasis on the *logos* provides, as it were, a philosophic justification for his ambiguous style.

Perhaps no other Greek author—or none except Aeschylus—has so systematically exploited the

possibilities of ambiguity and allusiveness that are implicit in all human speech. The characteristic expression of this ambiguity is in word-play, of which the fragments are full.<sup>14</sup> The most obvious example is that of the bow, whose name is life but whose work is death (fr. 48). Here the apparent opposition between word and deed, between name and reality, points in fact to their latent agreement: life and death, like day and night, are *one*; and the drawn bow is the symbol of their unity.<sup>15</sup>

The statement concerning the name of the bow must be read not only as a pun but as a riddle, whose meaning we can guess if we understand Heraclitus' view of the unity of life and death. There are other statements in the fragments, unintelligible at first sight, which can properly be understood only as riddles. Heraclitus himself refers to the riddling language of the Delphic oracle, which "neither speaks nor hides its meaning, but gives a sign (*sêmeinei*)" (fr. 93). The parallel suggests once more that Heraclitus' procedure is conceived of less as a literary device than as an inevitable consequence of human ignorance and the recondite nature of truth: like the utterance of Apollo, what Heraclitus has to say is necessarily enigmatic, because human beings do not have the "insights" which divine wisdom can take for granted (fr. 78). For the modern critic, these puns and riddles seem designed to dazzle the reader, not only to catch his attention but also, perhaps, like the Socratic "sting" of ignorance, to produce a temporary torpor of the mind that may prepare it for further insight. But Heraclitus would probably have refused to consider this a result of his skill in manipulating words. It is language itself which, by its dual capacity to reveal and obscure, provides the natural "sign" for the multifarious and largely latent connections between things. Like the Sibyl, who prophesies in virtue not of her own knowledge but that of the god, Heraclitus speaks to us not in his own name but as prophet of the *logos*, the rational "account" founded in the cosmic structure that makes all things one.

<sup>13</sup> For *koinos logos*, see the Herodotean usage mentioned in n. 11, and the passages listed in Powell's *Lexicon to Herodotus* (9 examples).

<sup>14</sup> For word-play, see Bruno Snell, "Die Sprache Heraklits," *Hermes*, vol. 61 (1962), esp. pp. 369-73. For the use of proportional formulae (of the type "A is to B as B is to C") see H. Fränkel, "A Thought Pattern in Heraclitus," *American Journal of Philology*, vol. 59 (1938), p. 309 (reprinted in *Wege und Formen*, pp. 253 ff.).

<sup>15</sup> There has been some fruitless debate about whether the bow of fr. 51 is drawn, or merely strung. There is no doubt that a bow "whose work is death" must eventually be more than strung; and I take it that the imagery of fr. 51 will likewise be more dynamic if the bow is drawn as well as strung, and the lyre played as well as tuned.

## III. "KOSMOS": FIRE AS THE ONE

An understanding of Heraclitus' views on cosmology is complicated by the fact that, whereas the doxographical tradition gives a rather detailed account of what purport to be his physical doctrines, only a relatively small number of the extant fragments are explicitly concerned with questions of natural philosophy. Furthermore, among the "physical" fragments are to be found such metaphorical statements as "All things are exchanged for fire and fire for all things," such ambiguous assertions as "The way up and the way down are one and the same," and such apparently banal remarks as "If there were no sun, it would be night." There has, therefore, been an understandable reluctance among modern scholars to accept the classical view of Heraclitus given by Aristotle and his followers, which makes of him an Ionian natural philosopher of the same type as Anaximander or Anaxagoras.

Even in antiquity there was perplexity on this point. The ancient *littérateur* who forged the correspondence between Heraclitus and Darius was no doubt drawing upon Hellenistic commentaries when he makes the Persian king say to our philosopher: "in some places, if your discourse is interpreted literally it seems to contain speculation concerning the entire world-order and what occurs within it . . . but there is much room for doubt, and the most learned are at a loss as to your true meaning."<sup>16</sup> At least one ancient critic was more positive; disputing the accuracy of its customary title, he maintained that the book was "not 'about Nature' but about man's life in society; what is said concerning Nature serves as paradigm or example."<sup>17</sup>

These ancient statements, based upon a familiarity with the book as a whole, confirm the impression made by the extant fragments that Heraclitus is not to be interpreted simply as a natural philosopher, a *physikos* in Aristotle's sense. From the persons whom he chooses to attack and (occasionally) to praise, it is clear that he thought of himself primarily as a rival of the poets and the

wise men. Not the scientists, Anaximander and Anaximenes, but the sages Thales and Bias, Xenophanes the poet, and Pythagoras the founder of an elite society are those whom he mentions.<sup>18</sup> Pythagoras and Xenophanes provoke particular scorn, for they have tried to expand the new philosophy of nature into a general view of god and man, and have (in his view) conspicuously failed. It is precisely this task which Heraclitus undertakes. His real subject is not the physical world but the human condition, which for the Greeks means the condition of mortality. His aim is to conceive the relationship of life and death within a universal doctrine of opposition, transformation, and hidden unity. Hence, when he talks about the transformation of the elements, he not only speaks in terms of their reciprocal birth and death; he directly identifies these processes with the birth and death of the soul (fr. 36).

Nevertheless, it is clear that Heraclitus' starting-point was the cosmology of Miletus. What he says about the transformation of the elements and the passage of opposites—hot, cold, dry, and wet—into one another is closely related to the thought of Anaximander's fragments, according to which all things perish back into that from which they have arisen. Heraclitus' mention of "justice" and "what is due" in fr. 80 seems actually to imply a reference to Anaximander's text. On the other hand, the elemental sequence of fire-water-earth in fr. 31 inevitably suggests the doctrine of Anaximenes:

When the air is spread out in finer form, it becomes fire; thickening, on the other hand, it forms winds; and clouds are produced by compression; water by further compression; thickened still more it becomes earth and, in its thickest form, stones.<sup>19</sup>

It is apparently this theory of continuous elemental change which inspired Heraclitus' description of the so-called cosmic flux: "it scatters and again gathers . . . forms and dissolves, arrives and departs."<sup>20</sup>

Now if Heraclitus has failed to mention air, winds, clouds, and stones in his own version of the cycle of transformation that is not, I think, because

<sup>16</sup> Diogenes Laertius, IX. 13.

<sup>17</sup> *Ibid.*, IX. 15, quoting an otherwise unknown critic named Diodotus.

<sup>18</sup> Vlastos interprets Heraclitus' silence concerning Anaximander and Anaximenes as a sign of respect ("On Heraclitus," *American Journal of Philology*, vol. 76 [1955], p. 356). In the case of Thales, however, there is nothing to suggest that the mention was hostile (fr. 38).

<sup>19</sup> Anaximenes A7.3 (Diels-Kranz).

<sup>20</sup> Fragment 91. For Heraclitus' debt to the Milesians, see further Vlastos, *op. cit.*, pp. 355-356 with n. 40. I assume that Heraclitus was familiar with a theory of "elemental flux" of the type suggested by the doxography for Anaximenes quoted above. For details, see my *Anaximander and the Origins of Greek Cosmology*, pp. 149-154.

he wished to propose the incredible three-element theory that has been devised for him by modern commentators. He abbreviates the Milesian list (i.e., account of the elemental transformation) because he has no intention of displaying his *polymathia*, that "learning of many things," which his predecessors mistook for wisdom. He can scarcely have meant to deny the importance of the atmosphere as one of the fundamental constituents of the universe; and in fact, its role is hinted at in his reference to the *préstér* or fiery whirlwind. The motive for introducing this rare and impressive form of atmospheric disturbance in place of the prosaic Milesian list (clouds, winds, rain, etc.) is surely the same as that which led him to replace the abstruse *apeiron* or *aér* by an "everliving fire." His intention is not to alter the Milesian cosmology in detail, but to reinterpret its total meaning by a radical shift in perspective.

Heraclitus' new conception is stated in the most arresting of his cosmological fragments:

This *kosmos* no one of gods or men has made, but it ever was and is and will be, an everliving fire, kindled in measures and in measures put out (fr. 30).

The Milesians were concerned to show how the order of the world had come into being, how it was maintained, and (very probably), how it would eventually perish, only to be produced anew out of its eternal and inexhaustible source. Anaximander had conceived this order as governed from without, by the primordial Boundless; Xenophanes had replaced the Boundless with an intelligent deity, who moves all things by thought. Heraclitus accepts the Milesian view of a world-order in which the opposition and transformation of elementary powers is governed by measure and proportion. But he denies that this order is imposed upon the world by any power from without. Instead, he deifies one of its internal constituents. For to say that fire is "everliving," that it "ever was and is and will be" is to say, simply, that it is eternal and divine. Yet Heraclitus insists upon the fact that this god participates in the changing life of nature, "kindled in measures and in measures put out." There is a genuine parallel here to Anaximenes' conception of the primordial Air. But Anaximenes would scarcely have emphasized the extinction of his principle at the very moment that he asserts its eternity, nor would he have identified his elemental principle with the cosmos

as such. It is one thing to say "all things are derived from air," or even "all things, in the last analysis, *are* air" (though the second formula is, I believe later and post-Parmenidean); it is something else to identify such an elemental principle with the concrete world-order it has engendered, particularly when the principle itself appears as a partial constituent of this order. Now fire, since Anaximander, was held to be a special constituent of the sun, stars, and, in general, the outer heavens. Thus Heraclitus' statement confronts us with the double paradox of a cosmic whole identified with one of its constituent parts, and an eternal principle embodied in the most transitory of visual phenomena.

The resolution of these antinomies is inseparable from that of other paradoxical statements, concerning the hidden "attunement" or fitting-together whose fastenings are "wholes and not wholes, convergent divergent, consonant dissonant, from many one and from one many" (fr. 10). The eternal life of the perpetually quenched fire, on the other hand, suggests still another set of paradoxical identifications: "Immortals are mortal, mortals immortal, living the others' death, dead in the others' life" (fr. 62). This network of interrelated paradoxes makes it clear that Heraclitus' doctrine of fire cannot be understood in physical terms alone.

In fact, the mere comparison with Anaximenes' theory shows that it was not for physical reasons that Heraclitus placed Fire rather than Air at the head of his list of transformations. The atmosphere is clearly the natural starting-point for a theory of condensation and rarefaction, since only in the movement from cloud-vapor to water and back again does one actually see such processes at work. The advantages of Fire are of a different sort altogether. They have been well described by Susanne Langer, in a passage that has no reference to Heraclitus:

Fire is a natural symbol of life and passion, though it is the one element in which nothing can actually live. Its mobility and flare, its heat and color, make it an irresistible symbol of all that is living, feeling and active.<sup>21</sup>

Heraclitus would certainly not have mentioned passion in this connection; instead he might have illustrated his meaning by pointing to the role of fire in the service of the gods and the cremation of

<sup>21</sup> Susanne K. Langer, *Philosophy in a New Key* (3rd ed., N.Y., 1957), p. 145.

the dead, or by alluding to the old stories in which a child is exposed to fire in order to absorb its superhuman life.<sup>22</sup> Yet Professor Langer's notion of a symbolic form provides us with the proper terminology for describing the status of fire in Heraclitus' thought. "Symbolical" is, after all, the modern way of describing a style which "neither declares nor hides but gives a sign" (*sêmeinai*). Thus for Heraclitus fire is neither a physical element nor a visual image *alone*. It is rather the concrete sign or symbol for a whole range of ideas, and above all for the series of interrelated paradoxes just cited. In fact, fire symbolizes not only the paradoxes, but their solution as well. For it represents life in death, identity in change, pattern in transience.

Now the specific character of a symbolical transformation (in the sense here employed) lies in the acquisition of new meanings without loss of the old. Heraclitus' fire is still that of the sacred altar flame, and also that of the Milesian theory of elemental change. And so his cosmos too is the world-order of Ionian physics, but profoundly altered by its new and wider dimensions. The periodic order of sun, stars, and seasons, with the related meteorological sequence of elemental change, has itself become the symbol for that cyclical community of life and death which forms the innermost content of Heraclitus' *logos*. Cosmology and astronomy interest Heraclitus not as special fields of knowledge but as contributions to *wisdom*: they provide the clearest insight into that universal reality which men encounter daily, but whose sense they fail to grasp, as the sleeper fails to grasp the significance of his actions in the realm of dreams.

This dual role of Fire, as element and symbol, provides us with one solution to the antinomies we have noted in the identification of fire and cosmos. By virtue of its concrete properties, fire serves as the embodiment of the cosmic flux between opposites, defined by rhythm and form. Like the river, it remains the same despite—or rather because of—the fact that new flames are ever pouring in. Its very existence consists in the passage from a point below, which is not burning, to a point above, where the flames have ceased. It emerges from the cold, produces intense heat, and vanishes once

more into its opposite. Hence this constituent part of the Universe may serve as the appropriate symbol for the rhythmic pattern of the whole. And since any new fire is, always and everywhere, recognizably the same as the old in essential properties (even if details of smell or color will differ), the agent of destruction may, by its own rebirth, symbolize the unceasing renewal of life. It is from this wider point of view, and not by reference to an element theory alone, that we must understand the statement of fr. 67:

The god is day and night, winter and summer, war and peace, satiety and hunger. It becomes other just as fire, when it is mingled with incenses, is named according to the fragrance of each one.<sup>23</sup>

It is clear that there is an underlying reference here to natural phenomena, but that the physical doctrine has been reinterpreted as the vehicle for a larger set of meanings.

Once we have grasped this symbolic transformation that Ionian cosmology has undergone at Heraclitus' hands, most of the controversial details of his scheme lose the crucial importance they have assumed in most modern discussions. Like the substitution of Fire for Air, any changes in detail must have been designed not to improve the physical scheme in a scientific sense, but to render its symbolical function more vivid and more drastic.

Thus for Heraclitus, "War is common and justice is strife, and all things take place according to strife and what is due" (fr. 80). Heraclitus' motive for introducing "justice" (*dikē*) and "what is due" or "needful" (*chreon*) in this connection must be explained historically, by reference to Anaximander. Anaximander had described the excess of any one elemental principle over its opposite in nature as a wrong-doing or injustice (*adikia*), for which it must atone (*dikēn didonai*). Such retribution, fixed by the "ordering of Time," takes place according to "what is due." Thus the order of periodic compensation is both necessary and just; and it is no distortion of Anaximander's language to say that he is proclaiming the rule of law in nature.

Heraclitus is in essential agreement with this

<sup>22</sup> See, e.g., the Homeric Hymn to Demeter, 237 ff., with the parallels cited by Allen and Halliday, *ad loc.*

<sup>23</sup> The word fire (*πῦρ*) has dropped out by haplography after *δύωμεν* in fr. 67, but it must be read or understood if the comparison is to make sense. Fränkel wished to read "oil" (*ἐλαιον*) instead of fire (*Transactions, American Philological Association*, vol. 69 [1938], pp. 230 ff.). But the loss of this word would be inexplicable; and the doctrine which results (and which is that of Plato in the *Timaeus*) bears more resemblance to Aristotle's *physikos* than to the Heraclitus of the fragments. The latter is not a "material monist" in the manner of Diogenes of Apollonia, with a physical theory based upon a simple underlying substance.

view, but he expresses it even more strongly and (from the Greek point of view) more paradoxically by refusing to recognize "injustice" at any point in the cosmic conflict of opposites.<sup>24</sup> Anaximander's image of justice is like that of Aeschylus in the *Oresteia*: the pendulum swings too far in one direction, and justice forces it back again in retribution. For Heraclitus, thrust and counter-thrust belong together in the common pattern, which is at once War and Justice, Strife and Compensation "according to what is due."<sup>25</sup>

Here again, Heraclitus is reformulating the Milesian view in the light of his principle of the total unity of things. Thus he insists upon seeing community above all at those points where contrast and discord are most evident, namely in the opposites. Each of his references to the harmony or unity of opposites exemplifies, from a different point of view, that common order in virtue of which all things are one. But this unity is never seen as purely rational and abstract, and in reinterpreting Anaximander's doctrine of cosmic justice Heraclitus has in fact endowed it with a richer human resonance and a more apocalyptic sense. When he states that, if the Sun should ever overstep his measures, "the Erinyes, ministers of Justice, will find him out" (fr. 94), the invocation of the Furies gives their mistress Dikê a mythical concreteness, and an immediate human significance in terms of crime and punishment, for which there is no parallel in the Milesian conception of cosmic justice. It is not merely that Heraclitus has a more poetic style than Anaximander. The form of speech points to that unity between the life of man and the order of nature which is at the center of all his thought.

The same may be said of the enigmatic fr. 120: "The bounds of Dawn and Evening are the Bear and, opposite the Bear, the Warder of bright Zeus." The "bounds" or "goals" of Dawn and Evening must be the points on the horizon which correspond to the limits of annual variation for

sunrise and sunset, what later authors call the summer and winter risings and settings.<sup>26</sup> These turning-points (for that is the normal sense of *termata*) mark the horizontal "measures" of the sun's course. How they are indicated by the Bear and the Warder is not clear, but I conjecture that Heraclitus is alluding to some ancient astronomical technique for plotting the sun's annual course among the stars. It has apparently not been remarked that the Warder (*ouros*) opposite the Bear (*arktos*) can only be the Bear-Warder (*arkto-ouros*), i.e., Arcturus in Boötes, a well-known star reference as early as Hesiod.<sup>27</sup> The precise astronomical link between these constellations and the mention of Dawn and Evening remains obscure, but the real meaning is perfectly clear: the faultless regularity of the sun's course is again expressed in mythic terms, by reference to the astral watchman appointed by Zeus to guard the polar constellation. If Zeus here is called *aithrios*, "Zeus of the bright sky," the title reflects the fact that, for Heraclitus, the ruling Wisdom of the universe is not to be identified with everyman's Zeus—not with the philandering, hen-pecked husband portrayed by Homer—but only with the cosmic power symbolized by Fire and by the thunderbolt which "pilots all things."

#### IV. "NOMOS": THE ONE IN THE CITY OF MEN

Developing, from a different point of view, his reinterpretation of Milesian cosmology as a paradigm for the order in men's own life, Heraclitus expounds a conservative and aristocratic political view, of which the historical details escape us, but whose fundamental principle is as fertile for the future as it is independent of any particular philosophic position. The conception of Law as the social and human expression of the rational order of Nature is perhaps the most influential single idea in the history of political theory. Accepted by Plato as the philosophic foundation for his last

<sup>24</sup> Compare fr. 102: in god's view, all things are fine and good and just, but men have taken some as unjust, others as just.

<sup>25</sup> See the comment of Vlastos (*American Journal of Philology* [1955], p. 358): "Two of the fundamental ideas in Anaximander—that there is strife among the elements, and that a just order is nevertheless preserved—are reasserted in a form which universalizes both of them and thereby resolves the opposition between them: what is a 'nevertheless' in Anaximander, becomes a 'because' in Heraclitus."

<sup>26</sup> See *ἀνατολή θέρων, δύση θέρων*, etc., in Aristotle, *Metæorologica* 363b4 ff., with the diagram provided by H. D. P. Lee in the Loeb edition, p. 187. Thus the riddle does refer to the cardinal points, but in the technical, astronomical sense which allows for seasonal variation. *Arktos* then may refer both to the North and to the constellation, but the second reference is the more significant (since it alone explains the *ouros*), and there is no allusion to the South. Heraclitus is concerned here with something more significant than the four points of the compass.

<sup>27</sup> For the seasonal importance of Arcturus in early Greek science, see the Hippocratic treatise *Airs, Waters, Places*, chap. 11 and *Epidemics I*, 4, 13, 17, etc.

and longest work,<sup>28</sup> developed by the Stoics in conscious imitation of Heraclitus, transmitted to the Roman and medieval world by Cicero's eloquent exposition in the *De Legibus*, the doctrine of natural law eventually gave birth to the principle of natural rights, and thus provided a justification for the American Declaration of Independence as well as for the French proclamation of the Rights of Man and Citizen.

In the course of this development, the concept of "law" has undergone important changes, above all under the influence of Roman and Hebraic notions of lawmaking. Our ideas of law tend to reflect either the divine Lawgiver of the Old Testament or the Roman conception of *lex* as a precise verbal formula. Neither notion is relevant to an understanding of what Heraclitus says about *nomos* in fragments 44, 33, and 114. The early Greek sense of *nomos* is that of a traditional order of custom and usage—normative, of course, and theoretically derived from the gods, but sanctioned above all by time and habit. Heraclitus wrote when this archaic conception of *nomos* was undergoing change in two directions. On the one hand, the new importance of written law codes, drawn up under democratic pressure, was giving rise to the view of law as a rational, man-made standard. On the other hand, increased knowledge of and interest in the traditions of other lands tended to give *nomos* (or more frequently, the plural, *nomoi*) the relative sense of "mores," the conventional pattern of life of a particular community. This interest in what we may call comparative sociology is clearly evidenced in Xenophanes' references to the religious customs of Thrace and Ethiopia, or in the contemporary researches of the world-traveler Hecataeus. It is easy to see how such comparisons between different cultures would tend to erode the traditional view of *nomos* as god-given and sacrosanct. These two trends, sociological relativism and the reformist tendency to regard law as man-made, converge in the late fifth century in the classic antithesis between *physis* and *nomos*, Nature and Convention.

Heraclitus is faced not with the antithesis but with the trends that produced it; and his own reaction is to resist them both, in the name of the

divine unity of the cosmos. More exactly, he resists democracy as such, but cultural relativism only insofar as it tends to deny the ultimate validity of law. His notion that the *nomoi* of men, different as they may be, are all "nourished by the divine" (fr. 114), has its roots in the old mythic view which sees society, nature, and the gods as only different aspects of a single, undivided realm. The philosophical achievement of Heraclitus is to have preserved this mythic sense of coherence between men and nature while giving it a rational form. He does so by linking the civic community to the universal principle of what is "common" in nature, thus relating the pattern of *nomos* to the rational structure of the *logos* and to reason itself (*nous*). Like Solon, Heraclitus sees that the moral existence of the civic community depends upon a shared framework of custom and law, just as its physical existence depends upon the defense of its city wall. Hence *stasis*, or civil war, like *hybris*, or individual violence (fr. 43), is recognized as the disaster which destroys the civilized life of man.<sup>29</sup> Heraclitus goes beyond Solon, however, in grounding this sociological justification of law and order in a philosophic principle which is fundamental and universal: rational comprehension (*nous*) is the knowledge of what is common (*xunos*); hence the *nomos* of the human community is a direct expression of the "divine," that is, of the rational unity of the world. Not only the individual man but society too is a microcosm, a reduced model of the world-order. Like the soul, the city has a structure "of which one cannot find the limits," since it draws its strength and nourishment from the eternal structure of the universe.

## V. PSYCHOLOGY AND THEOLOGY

For Heraclitus more than for most thinkers, the question of deity is inseparable from that of the human soul: "Immortals are mortal, mortals immortal, living the others' death, dead in the others' life" (fr. 62). We shall see that this paradox points to what is most essential in Heraclitus' view of the life of man and god.

Heraclitus had inherited from his Ionian predecessors a physical view of the *psyché* as the

<sup>28</sup> See Plato, *Laws* IV, 714a, where the play on *nomos* and *nous* echoes Heraclitus fr. 114. Book X of the *Laws*, like the *Timaeus* as a whole, is essentially concerned with the reinterpretation of Nature in terms of Reason—a reinterpretation different in kind, but not in tendency, from that of Heraclitus himself.

<sup>29</sup> See Solon's elegy (Diehl, fr. 3) for a definitive statement of the archaic conception of *stasis*, *hybris*, and *sunomia*. Solon's description of public disorder leaping over the walls of private houses in which men try to hide may have inspired Heraclitus' comparison of *nomos* to the city wall.

life-breath—the power of breath within us, as well as the concrete air we breathe. That is what Anaximenes must have meant by identifying the soul with “air.” Heraclitus too sometimes thinks of the *psychê* in purely physical or elemental terms, as we see from fr. 36:

For souls it is death to become water; for water it is death to become earth; out of earth water arises, out of water, soul.

What does this imply for the physical constitution of the *psychê*? It is commonly held that when Heraclitus says “soul” here we must understand him to mean fire. But such a view can be justified only by the modern assumption that Heraclitus recognizes only three elements: earth, water, and fire.<sup>30</sup> If we ignore this assumption, what we expect to arise from water is not fire, but vapor, steam, or “air.” That Heraclitus means us to understand something of the sort is clear not only from the parallel to Anaximenes’ doctrine of the soul, but from the ancient testimony concerning Heraclitus’ own view, which uniformly states that the soul is to be explained as an exhalation (*anathymîsis*) or breath (*pneuma*). Furthermore, this view is decisively confirmed by the fragments that describe the *psychê* as becoming either moist or dry (frs. 77, 117–18). Such statements make sense for a breath or exhalation, but not for a fire.

Among the ancient sources which report Heraclitus’ view of the soul as a vapor or exhalation, the most interesting is an indirect quotation from Cleanthes, who links this view to the image of the cosmic river:

Heraclitus likened souls to rivers when he said, “As men step into the same river, other and other waters flow upon them.” And souls too are exhaled from moisture. (fr. 12)

Although Heraclitus’ river is usually interpreted as a general reference to elemental transformation, this ancient author believes that it refers to the soul. In a sense, he must be right. For if Heraclitus

were referring solely to the elements, what would be the point of the men stepping in? This intrusion of the human participant gives the notion of identity-in-flux an immediate, concrete significance.<sup>31</sup> If our soul is an exhalation, it too is an on-flowing process, and it may well be that the river into which we cannot step twice is the psychophysical flux of our own lives and experience.

In a larger sense, the river certainly refers to changes throughout the universe. But the notion of the soul as an exhalation from moisture also links it to the larger cycle of elemental transformation, to the “upward path” from earth to water to vapor, and eventually on to fire.<sup>32</sup> This is the road of life for the soul; the road downwards is that of death (as is explicitly stated in fr. 36). For Heraclitus, the identity of life and death must be understood in intimate connection with the identity of the way up and the way down. In physical language, the way up is that of the *dry* souls, which are drawing nearer to fire; the way down is that of the moist, who are perishing into water. In moral and psychological terms, the way down is the life of pleasure, passion, drunkenness, and dissolution (frs. 77, 117; cf. fr. 85); the way up is that by which souls become “best and wisest” (fr. 118). With his characteristic rigor, Heraclitus identifies the conflicting human tendencies toward wisdom and sensual pleasure with the opposing courses of elemental transformation.

On this view, an unimaginable surprise is in store for men when they die. What will happen to those who “sate themselves like cattle” is clear enough (fr. 29): they simply pass into the inert elements, water and earth.<sup>33</sup> But greater fates are reserved for greater deaths (fr. 25), and the nobler destiny certainly connotes a movement upwards: to smoke, perhaps,<sup>34</sup> and ultimately to fire. It can only be these nobler souls, the kinsmen of Fire, “who rise up and become wakeful watchmen of the living and dead”—the task which Hesiod assigned to the Golden Race of mortals after their

<sup>30</sup> Contrast the ancient view, which inserts air between water and fire in every version of fr. 76. Kirk (*Heraclitus*, p. 343) strangely uses the assumption that *psychê* in fr. 36 must mean fire as *proof* that Heraclitus believed in only three elements, whereas his treatment of fr. 36 takes for granted the proposition he wishes to prove.

<sup>31</sup> A significance which is blunted by Cratylus’ insistence that we cannot even step in once, since we are changing too. He thus destroys the paradox of *identity* in flux and flatly reduces the human aspect to another physical event, which Heraclitus does not.

<sup>32</sup> Compare Aristotle’s description of the cycle of evaporation and condensation as “a river flowing in a circle up and down, common to air and to water” (*Meteor*, 347a2).

<sup>33</sup> Compare Menelaus’ curse on the Greek heroes in a moment of cowardice: “May you all become water and earth!” (*Iliad* 7. 99).

<sup>34</sup> This seems to be the meaning of the enigmatic fr. 98: the (better) souls smell after death, because their next stage upwards is that of the purely *dry* exhalation, i.e., smoke (cf. fr. 7).

death.<sup>35</sup> The mass of mankind clearly has a lower fate: they wish only to live and die, or, what is the same thing, to beget children; this is the life and death of rest and sleep, the polar opposite of watchful waking.<sup>36</sup>

Thus a man's character is his *daimôn* (fr. 119), not only his deity but his divine guardian, his fate in this world and the next. It is simply and directly, by their passage back and forth into the elements, that "mortals are immortal, immortals mortal, living the others' death and dead in their life." "Life-time" (*aiôn*) is the royal player, moving the pieces back and forth in a game whose rules are those of the cosmos. It is War, the struggle of opposing tendencies up and down, which assigns everyone his place: god or man, slave or free. The fragment that asserts this supreme role of War, as "father of all and king of all" (fr. 53), is the exact inverse of fragment 30 on the world-order as everliving fire: this cosmos is made by neither man nor god; rather, it is the cosmic struggle itself that makes both gods and men.<sup>37</sup>

We note that the term "god" is thus used by Heraclitus in two senses. First, there are the "immortals" in the plural, who, like the mortals with whom they exchange roles, are only stations or stages in the cosmic course—the so-called elements and world-constituents of every kind. But there is also the "god" in the singular, who is compared to fire and equated with war and peace (fr. 67), who is "the one divine" source of all human laws (fr. 114), the thunderbolt who pilots all (fr. 64), the "one wise alone" who is and is not the same as Zeus (fr. 32). There is a suggestive parallel between this heno-polytheism and the position of Xenophanes, whose "one god" is "the greatest among gods and men."<sup>38</sup> But as soon as this distinction between many "gods" and the one "god" has been drawn, its limitations must be recognized. For Heraclitus, all things are *one*, and the unity which includes Night and Day also comprehends mortals, immortals, and the one wise alone, even if the latter is in some sense "set

apart from all else" (fr. 108). It is set apart, not only because it is of a higher logical type, the whole which comprehends the parts, but also because it dominates the rest. The unity of wisdom consists in the insight "how all things are steered through all things" (fr. 41). The exact wording of this fragment is uncertain, but the sense surely is that cosmic guidance is a function of cosmic intelligence. Yet the statement seems to be intentionally ambiguous, and to point both to the uniqueness of the supreme Wisdom, and at the same time to the possibility of human insight into the cosmic pattern. This ambiguity is implicit in the term *sophos* itself, which can denote technical competence, practical wisdom, theoretic knowledge, and divine omniscience. This ambiguity is artfully exploited by Heraclitus in a series of ambivalent assertions.

It is wise, listening not to me but to the *logos*, to agree that all things are one. (fr. 50)

Of all those whose *logoi* I have heard, no one has come so far as this, to recognize that the wise is set apart from all else. (fr. 108)

The wise is one, to master the insight how all things are steered through all things. (fr. 41)

The wise is one alone, willing and unwilling to be called by the name of Zeus. (fr. 32)

It is clear that in the first of these four statements, wisdom is a quality expected of Heraclitus' audience, while in the fourth case it is identified with the unique divine principle. The second and third statements are ambiguous, and the commentators disagree as to whether they should be taken as applying to the wisdom of man or god. But such debate is pointless, since Heraclitus has been at some pains to make this strict dichotomy impossible. As we have seen, even in the first of these four assertions (fr. 50), the wisdom of the intelligent hearer consists in an agreement with the *logos*, that is, in a congruence with the rational unity of things. The distinction between the knowing subject and the known object, like that

<sup>35</sup> Fr. 63; compare Hesiod, *Works and Days*, 121-123.

<sup>36</sup> Obscure as is fr. 20, it probably involves a reference to the 30-year cycle of grandfather-to-grandson (cf. Diels A 19); this is apparently Heraclitus' version of the mystic "circle of generation" in which most men are trapped.

<sup>37</sup> Why is this order called *aiôn*, "life-time," in fr. 52? The original meaning of the term is "vitality, vital force" (see E. Benveniste in *Bulletin de la Société de Linguistique*, vol. 38 [1937], p. 109); yet in Heraclitus' day it is frequently used in the sense of "duration, life-time," and almost as a synonym for time itself (*chronos*). I suspect that Heraclitus has chosen the word *aiôn* precisely as a cryptic equivalent for *chronos*, thus referring overtly to the alternation of human life and death, but covertly to the "ordering of time" which Anaximander had shown to be the pattern of the universe. This word-play would reflect the fact that, for Heraclitus, the two orders are literally one. Precisely the same ambiguity attaches to War in fr. 53: the cosmic struggle between opposites is represented in human terms by death in battle (fr. 24). Both human and cosmic war tend toward the same results, to send the noble upwards and the base below.

<sup>38</sup> See Xenophanes fr. 23 (Diels-Kranz).

between man and god, has only a relative, not an absolute significance for Heraclitus. "A man is found foolish by a deity as a child by a man" (fr. 79). Yet the consequences of such relativism are neither arbitrary nor irrational. Here as elsewhere there is a *logos*, a ratio, linking the opposed terms: man is to god as child (or ape, in frs. 82-83) is to man. The respective parts of wisdom or folly assigned to child, man, and god represent proportionate shares in a single reality. And this gradient scale of knowledge, which reaches from the zero point of total darkness and sleep to the maximum intensity of divine insight, coincides with the scale of being itself, that stretches from the inert nadir of earth to the ardent zenith of the everliving flame.

Another name for the unifying order of the world is *Dikē*, and we have seen how Heraclitus adapts Anaximander's concept of cosmic justice to bring out the older sense of *dikē* as a concrete principle, directly applicable in the affairs of men. "Dikē will lay hands on false witnesses and forgers of lies" (fr. 28). In so doing, she acts as an exact parallel to the oncoming Fire which "will judge and lay hands on all things" (fr. 66); for "how will one hide from that which never sets?" (fr. 16). Fire here takes over the traditional mythic function of the "all-seeing sun," while the epithet "oncoming" suggests a parallel to the on-flowing waters of the river.<sup>39</sup> In fact the judgment of Fire, and the law-enforcement of *Dikē*, are both identical with the cosmic process itself: the goddess who punishes liars is the same as she who regulates the course of the sun (fr. 94). Our Christian sources naturally interpret the judgment of Fire by reference to the end of the world, but such was not Heraclitus' meaning, even if (as I believe) he conceived of a periodic reabsorption of all things into the everliving Fire. For at that point the fate of hero and coward would be exactly the same. The justice of the world-order is rather a continuous, immanent aspect of the cosmic process throughout. It is identical with the movement of all things up and down.

The soul, as a spirit or vaporous exhalation, participates by its own nature in the movement upwards to Fire. The rare man with insight, who understands "each thing according to its nature and as it is" (fr. 1), will identify himself with this spontaneous tendency, for it constitutes his real

self. If by his life and deeds he has chosen Fire as his *daimōn*, it will lead him, like a faithful psychopomp, to its abode in the house of the dead, which in this case is the realm of increased life. The path of most men, on the other hand, who sell their souls for pleasure and passion, is the "way down" to moisture and extinction (fr. 15).

Thus, Heraclitus is a mystic, perhaps, but his eschatology involves none of the dualism that is so conspicuous in Empedocles. For the latter, as a true "Orphic," the world of elemental change is a hostile abode; the incarnation of the deathless *daimōn* means the putting-on of an alien garment of flesh. In such a view, the cycle of transmigration is necessarily limited to the forms of living things. Heraclitus remains truer after all to the spirit of Ionian physics than to the salvation cults of the West, and for him transmigration is of a different type altogether. It is not as creatures of a higher order that men participate in this mystic cycle; their death and rebirth is that of the elements and of the cosmos as well. The "greater portions" allotted to greater souls lie within the mundane order, not beyond it. The human soul transmigrates in the same way as earth and water.

Yet the way of the soul is after all not quite the same. Earth and water presumably have no choice as to their course of transformation. In the case of man, however, the tone of Heraclitus' entire discourse presupposes a dimension of moral insight and choice, of wisdom and folly, or, as a modern would say, of inner freedom. Like Spinoza, and like the Stoics, Heraclitus assumes that the principle of moral autonomy is compatible with a rigorously monistic scheme. But beyond the Stoics, and beyond nearly all thinkers of the classical age, Heraclitus is acutely aware of the specific character of psychic reality as such, with its characteristic dimension of "inwardness." It would be difficult to find another philosopher before late antiquity who could say, "I sought myself: I went in search of my *self*" (fr. 101). It is this new and rare sense for the human soul as an inner world, or as containing the secret of the universe within itself, which Heraclitus expresses in his denial that one can "find the ends of the soul by going, even if you travel over every way; so deep is its *logos*."<sup>40</sup> A similar thought seems to be expressed in his mention of the *logos* of the soul

<sup>39</sup> Compare *πῶρ ἐπελθόν*, in fr. 66 with *ἕδρα ἐνυππεί* in fr. 12.

<sup>40</sup> Fr. 45. See the excellent remarks of Bruno Snell, who suggests that the metaphor of depth (in *βαθὺς λόγος*) serves precisely to designate what is characteristically psychic, the soul's own dimension which is not that of mere spatial extension (*The Discovery of the Mind*, chap. 1, transl. by T. G. Rosenmayer, pp. 17 ff.).

"which increases itself" (fr. 115)—and which presumably might go on increasing until it merged itself in the total *logos*, in the life and death of the universe. Heraclitus appears to be struggling, for the first time in history, with the paradox of Pascal: that man is a thinking reed, so frail that he may be easily crushed, yet lifted above the universe by his own consciousness of his weakness and mortality. "Par l'espace, l'univers me comprend et m'engloutit comme un point; par la pensée, je le comprends."<sup>41</sup> Probably no philosophy has fully resolved this paradox, and certainly Heraclitus would not have been satisfied with a neat Cartesian dualism. The novelty, but above all the intrinsic difficulty of the problem he was confronting, accounts for his intense sentiment of the *hidden* character of the truth he proclaims.

There is no doubt that, in intent, his monism is as thoroughgoing as he could make it. It extends not only to the general unity of opposites, and to the fusion of mystic eschatology with elemental physics, but also to the identity of thinking subject and physical object within the divine unity itself. But here we must distinguish between the real unity which Heraclitus is trying to describe, and the specific identity of the different symbols by which he refers to it. It is too easy, and I think profoundly misleading, to say that god, Fire, the *logos*, War, Dikê, and "the one wise" are equivalent terms for Heraclitus. Each one is rather a "sign" that points to one aspect of an entity for which there is no single adequate formula, and which Heraclitus' whole discourse strains to express by every device at his command. The fitting-together which he is asserting throughout is essentially *hidden*, and therefore only riddle, paradox, and allusive word-play can do it justice. To speak plainly about such a subject would be to falsify it in the telling; hence Heraclitus is consciously and unavoidably "obscure." And similarly, it would be a betrayal on the part of the interpreter to force Heraclitus' language and imagery out of its bewildering complexity into any simple and straightforward doctrine. The most we can do is to define the respective range of the various symbols he employs; and that is what has been attempted here for Fire, War, Justice, and the upward and downward path.

When I say, therefore, that in Heraclitus' conception of the divine unity of things, the knowing subject and the physical object coalesce, I mean that his use of the term "wise" (*sophon*) in this connection implies that the cosmic order

symbolized by Fire and War is thought of as in some sense *aware* of its own structure, and that this awareness is in turn comparable to the human insight which Heraclitus' own *logos* is meant to convey. What "wisdom-loving men" (*philosophoi andres* in fr. 35) are ultimately seeking is the total insight of deity. The term *logos*, on the other hand, does not denote this insight as such but rather the structure which it apprehends, according to which all things come to pass—a structure of rational measure and proportion, expressible in discourse. Fire is the visible symbol of this order, of which "the wise" is in turn the rational apprehension. "God" (*theos*) for Heraclitus is the least precise term of the series; it is the final predicate of supremacy bestowed upon a nature defined by the other terms. It points to the total unity underlying *logos*, Fire, wisdom, and all other aspects of the divine whole, including the two complementary notions of Justice and Warfare.

Thus the power of Heraclitus' synthesis resides not only in the rigorous simplicity of his monism, but also in the intricate complexity of the web of symbols by which, like Apollo, he manifests his meaning. It is not only the artistry with which they are articulated, but the depth and suggestiveness of the "signs" themselves that make Heraclitus so extraordinarily relevant to the intellectual trends of our own age. We may therefore close with one specific sign or image whose use by Heraclitus makes him almost our contemporary, while it reveals at the same time the inner unity of his own thought: I mean his references to sleep and dreaming.

Other philosophers, like Descartes, or like Plato in the *Theaetetus*, have cited the experience of an alien world which opens up before the dreamer in order to raise the question of the reality of what surrounds us in our waking life. Primitive thought everywhere recognizes the dream experience as a possible road of access to the realm of the divine, or of the dead. Heraclitus invokes both of these attitudes, the awe of the primitive as well as the scorn of the epistemologist, but he does so with a multiple resonance that can scarcely be matched except by a modern poet or philosophic novelist:

Tout à coup je m'endormais, je tombais dans ce sommeil lourd où se dévoilent pour nous le retour à la jeunesse, la reprise des années passées, des sentiments perdus, la désincarnation, la transmigration des âmes, l'évocation des morts, les illusions de la folie, la régression vers les règnes les plus élémentaires de la nature (car on dit que nous voyons souvent des animaux en rêve, mais on oublie que presque toujours

<sup>41</sup> *Pensées*, nos. 264-265.

nous y sommes nous-même un animal, privé de cette raison qui projette sur les choses une clarté de certitude; nous n'y offrons au contraire au spectacle de la vie qu'une vision douteuse et à chaque minute anéantie par l'oubli, . . .) tous ces mystères que nous croyons ne pas connaître et auxquels nous sommes en réalité initiés presque toutes les nuits, ainsi qu'à l'autre grand mystère de l'anéantissement et de la résurrection.<sup>42</sup>

It is something of this mysterious, rich, and imprecise revelation of the dream world that Heraclitus evokes. Such a significance is only latent when, at the very beginning of his work, Heraclitus compares the waking thoughtlessness of most men to the oblivion of the sleeper (fr. 1), or the narrow vision of the crowd to the private world of slumber, both of them deprived of that universal character which is the stamp of reason and reality (frs. 89, 73, 2, and 17). But even in sleep one does not escape from the cosmos (fr. 75). Since all things are one, since the unity of the *logos* is all-pervasive, the dream-world too is an aspect of this universal order. It is simply the region where Fire is most nearly quenched, where the link with the Common is particularly unstable.

Death is everything we see awake; all we see asleep is sleep (fr. 21).

A man strikes a light for himself in the night, when his sight is quenched. The living grasps the dead asleep; the waking grasps the sleeping (fr. 26).

If what we see in the waking world is death—the death of Fire, presumably, as well as of soul which has passed into the elements—the sleeping world does not present us with death's reciprocal, life, as the parallelism of the clauses in fr. 21 would lead us to expect. On the contrary, the sleeper enters into an even more intimate contact with the realm of the dead, as dreams have always testified; in place of the light of vision, another lamp is kindled for the man who sleeps, an illumination which can only be that of the soul itself (fr. 26). The soul is the link by which the waking grasps the

sleeping, the sleeping grasps the dead. Thus the dream experience is one more revelation of the endlessness of the soul: one cannot find its limits; for there is no region in death or in life to which the soul does not travel, waking or asleep.<sup>43</sup>

In a sense, then, the soul also travels upward, to the Wisdom which is "set apart from all else." Yet full insight into the steering of all things belongs to god, and not to man: it is the thunderbolt of Zeus which pilots all, and the wisdom of the steering is Zeus himself. Neither sleeping nor waking, neither erudition nor science can bring a human being to this state: as a man, he will not be wise, but only a lover-of-wisdom, a *philo-sophos*. It is to this ardent search for insight into the divine unity of all things that the words of Heraclitus would summon the reader. It is such *philosophia* which constitutes for him true piety.

And it is against such an insight that traditional piety is to be measured. It does not come off well. Even the more austere temples and images reflect a gross ignorance of the true nature of deity, as Xenophanes had seen. But the orgiastic rites of the mysteries and the phallic cult and maenadism of Dionysus are the reverse of holy: it is not drunkenness and frenzy which can prepare the soul for insight. Yet even in such superstition the knowing eye discerns the fabric of cosmic unity. The madness and obscenity of Dionysiac worship are after all more appropriate than one would suppose, for Dionysus is the same as Hades (fr. 15): the apotheosis of drink and sensuality is in reality the god of death, the *daimôn* of the downward path of the soul to passion, pleasure, and dissolution. It is therefore only fitting to honor the god of destruction with the symbols of debauch. Justice will lay hands on false witnesses and forgers of lies, not in some mythical scene of judgment but in the natural order of things, in the relentless course through world and elements of the everlasting fire, the sun that never sets.

Colombia University

<sup>42</sup> Marcel Proust, *A la Recherche du Temps perdu*, Édition de la Pléiade, vol. I (Paris, 1954), pp. 819–20. Compare Proust's opening remark: "Une homme qui dort tient en cercle autour de lui le fil des heures, l'ordre des années et des mondes" (*ibid.*, p. 5). I see no reason to think that Proust was influenced by Heraclitus, but he also was one who sought in himself for the secrets that most men look for outside.

<sup>43</sup> I do not understand how critical scholars, including some who are normally most sceptical of secondary testimony, can seriously regard the report of Sextus Empiricus (Heraclitus A 16) as a valuable source of information concerning Heraclitus' theory of sleep. What Sextus gives us is nothing more nor less than a Hellenistic-Stoic exegesis of a few well-known Heraclitean statements in terms of later physical and psychological doctrines. To assume that, because the commentary is internally consistent and contains one very vivid image (namely, that of the coals which glow when brought close to the fire), it must therefore be based upon lost statements of Heraclitus, is to forget that some Hellenistic authors (such as Posidonius) were capable not only of intelligent commentary but also of powerful effects of style. (For a critique of the doctrines reported by Sextus see K. von Fritz, "voûs, voeiv, and their Derivatives," *Classical Philology*, vol. 40, [1945], pp. 234–35.)

### III. RECENT WORK IN METAPHYSICS

ROBERT C. COBURN

THE purpose of this paper is to draw attention to some of the more valuable and/or interesting contributions to the literature of metaphysics which have been made in the last few years. No attempt will be made, however, either to provide a truly comprehensive survey of recent work in metaphysics, or even to mention all of the contributions made in recent years which satisfy the vague description "valuable and/or interesting." There is too little space, too much recent metaphysics, and too large a gap in the writer's knowledge for either attempt to be sensible. However, I hope, despite these limitations, that what follows may be of some value to those whose philosophical studies have focused primarily upon problem-areas more or less removed from those traditionally allotted to metaphysics. It will be convenient to group the contributions I shall take note of under the following rubrics: Universals, Existence, Causality, Time, Persons, Identity, and God.

#### I. UNIVERSALS

Under this heading I wish to mention, first, several recent attempts to criticize and explain the attraction of certain notions which have commonly led to realistic doctrines; second, several new and spirited defenses of realism; and third, the treatment given by one philosopher of some refreshingly original questions concerning the very distinction round which realist-nominalist controversies have turned, *viz.*, the particular-universal distinction.

Among the notions which have often lain behind espousals of realism in one form or another are these: first, that a necessary condition of our being able to refer to a thing is that that thing exist; and second, that a necessary condition of there being an "objective justification" for the application of a general term is that its instances share something in common beyond the mere fact that they are its instances. One of the better recent

discussions of the former notion is contained in N. Wolterstorff's essay, "Referring and Existing" (*Philosophical Quarterly*, vol. 11 [1961]). In this essay Wolterstorff, among other things, (a) argues by appeal to statements like "The site of Atlantis has been sought by many" that the connection so often alleged to obtain between referring and existing simply does not hold and, more importantly, (b) sets forth considerations of a kind designed to explain and thereby dissolve the ability of this (as he thinks) patent falsehood to masquerade as an obvious truth. These considerations include a number of telling points about such expressions as "There is (are)," "something," "nothing," and "anything"—e.g., that the statement "There are things that don't exist" is not self-contradictory—as well as a discussion of some of the analogies and disanalogies which hold among various of the transitive verbs we use in describing (imputing, etc.) human actions. A particularly clear and persuasive, though avowedly unoriginal, discussion of the other notion mentioned as one of the pillars of realism is found in R. Bambrough's paper, "Universals and Family Resemblances" (*Proceedings of the Aristotelian Society*, vol. 61 [1960-61]). Here the emphasis, however, is upon criticism of the notion in question, the central point in this criticism being the now familiar one that the instances to which a general term applies form a "family" of a certain kind—a family which constitutes an open class, a class each of the members of which bears some "striking" resemblance(s) to at least some of its other members, and a class such that it is possible for human beings to learn how to identify its members by contact with only a few.<sup>1</sup>

Among the more impressive of the recent metaphysical essays in which realism is espoused are A. Donagan's "Universals and Metaphysical Realism" (*Monist*, vol. 57 [1963]), and D. S. Shwayder's *Modes of Referring and the Problem of Universals* (University of California Publications in

<sup>1</sup> Mention should perhaps be made in passing of the vigorous attempt recently made by L. J. Cohen in *The Diversity of Meaning* (London, Methuen & Co., 1962), chap. 4 to understand certain of the historical controversies concerning the existence of universals in a way which suggests the irrelevance of such discussions as the above.

*Philosophy*, vol. 35 [1961]). The version of realism for which Donagan provides a partial defense, however, will doubtless strike many as a rather puny and unexciting one next to (say) Plato's—indeed, one hardly needing “defense” at all. For what Donagan's realism amounts to is merely the denial that all predicates can be dispensed with in a true account of the nature of things, together with the assertion that the truth of a description depends upon the way the world goes. Shwayder's realism is, by contrast, considerably more exciting; however, what his doctrine gains in excitement, it loses in plausibility. Roughly speaking, it is Shwayder's view that to say the color blue or the number 8 exists is to say neither more nor less that there exists a certain use of (a) certain (family of) expressions; in other words, for Shwayder such universals “reduce to” uses of expressions. (The “roughly speaking” is necessary in view of certain qualifications Shwayder introduces *passim*—see, e.g., *ibid.*, p. 48, n. 5.) That this view lacks plausibility becomes evident, of course, as soon as one begins to reflect upon the consequences it seems to carry. One of these, for example, would appear to be that such questions as “How long have the natural numbers been in existence?” and “What do you think the chances are that 3 and 8 will cease to exist by next March?” make perfectly good sense inasmuch as “uses” of expressions are the sorts of things, one naturally supposes, that can (and do) come and go. But even those of the most latitudinarian semantic views will be sceptical of the legitimacy of these questions.

Doubtless the most important full-scale defense of realism is, however, that presented in W. V. O. Quine's *Word and Object* (New York and London, 1960). Quine's realism consists essentially in a reluctant admission that though many types of abstract objects can be abjured, at least one type cannot—namely, classes. One way of putting the argument which leads to this admission is the following. Many of the things we want to say about “reality” are best rendered, for a variety of reasons, in what Quine speaks of as “the canonical notation of quantification,” and in particular, by sentences of the form “ $(\exists x) (\dots x \dots)$ ,” where the variable ranges over a universe comprised (in part) of classes. But now if we utter or inscribe a sentence of this latter form with the intention of making a true statement, then we *eo ipso* commit ourselves to holding that those entities over which the bound variable ranges exist, and exist in the one, univocal sense this word carries in ordinary usage (p. 131).

Hence the inescapability of the version of realism indicated. Exactly why we should, according to Quine, accept the premisses of the argument when it is set out in this way is, of course, a long story, and, as regards the second of these premisses, a somewhat obscure story—for both of which reasons I shall make no attempt to recount it here.

The quite original questions concerning the particular-universal distinction to which I alluded above are raised by P. F. Strawson in his recently published *Individuals: An Essay in Descriptive Metaphysics* (London, 1959). Among the most central of these is the question as to what fact or facts lie behind the traditional doctrines (a) that while universals can enter into propositions as either subjects or predicates, particulars can enter as subjects only, and (b) that particulars alone exist—or anyhow exist in the primary sense of the word. The full answer which Strawson provides to the question as to the “rationale” (his word) of the doctrine concerning the non-predicability of particulars is too complex for adequate summary here. But a part of what he has to say may be noted by way of indicating something of his general approach—though even here I shall be expressing at best only the gist of his remarks. One way of elucidating the idea of the subject-expression of a subject-predicate statement (and *a fortiori* the notion of a logical subject), is, he points out, by characterizing such expressions negatively as those which do not introduce terms—where “terms” are here thought of as non-linguistic items—in such a way that what they are introduced into linguistically is a proposition or a propositional clause. It is then noted that thusly characterized subject-expressions are “complete” in a way in which those expressions are not which do carry linguistic indications that their terms are being introduced into propositions or propositional clauses, expressions such as “is a smoker.” Strawson next points out that particulars, however, are also “complete” in a way in which non-particulars are not. They are “complete” in the sense that particular-introducing expressions necessarily “carry a weight of fact” in virtue of its being the case that a necessary condition of a speaker's making an “identifying reference” to a particular, as opposed to a non-particular, is that he know an empirical fact of an individuating sort about the particular in question. Now in view of all of this, Strawson finally suggests, it is quite natural that particulars should come to be thought of as entering into propositions only via subject-

expressions since the "completeness" of both constitutes a sort of natural alliance between them. Strawson's treatment of the question concerning the rationale of the doctrine which accords a special position to particulars as regards existence, though considerably less complex than his treatment of the first doctrine, similarly turns upon his view that particular-introducing expressions "carry a weight of fact" in the sense just indicated. Here his suggestion seems to be that what lies behind the doctrine in question is just this fact about particulars together with the (undoubtedly deeply rooted) tendency to "wed the notion of existence to empirical fact."

## II. EXISTENCE

In addition to some of the writings mentioned above, the contributions made in recent years on this second topic that are of special interest include (1) several discussions which, by making clear something of the variety of speech acts which we perform in uttering sentences of the forms: "A exists," "There exist F's," etc., have shed light on certain traditional questions which turn upon the character of the concept of existence; (2) several other discussions which, though not directed to, clearly bear upon the question as to the cogency of the most fashionable type of existence-argument employed in recent years, *viz.*, the so-called Paradigm Case Argument; and (3) an extended consideration of the question as to the legitimacy of the existence-claims which are often made concerning the entities to which ostensible reference is made by various scientific theories.

The discussions I have in mind under (1) are those of P. Geach in *Three Philosophers* (Oxford, 1961) and *Reference and Generality: An Examination of Some Medieval and Modern Theories* (Ithaca, N.Y., 1962); W. Alston in "The Ontological Argument Revisited" (*Philosophical Review*, vol. 69 [1960]); J. Shaffer in "Existence, Predication, and the Ontological Argument" (*Mind*, vol. 71 [1962]); and K. Baier in "Existence" (*Proceedings of the Aristotelian Society*, vol. 61 [1960-61]). Geach points out that what we sometimes do in uttering sentence forms like the above are such things as: (a) assert of a concept that it has application or, alternatively, say of a general term that it is truly predicable of something, (b) assert of a name that it does in fact name an object, and (c) assert (or

imply that it is true) of an individual that it existed at or during some past time, and/or that it exists now, and/or that it will exist at or during some future time.<sup>2</sup> Some obvious examples of sentences the uttering of which under certain circumstances would be to perform acts appropriately described in these ways are: (a) "There exist winged horses," (b) "John Doakes really does exist," and (c) "The Boston Symphony Orchestra still exists." Alston suggests that in uttering sentences of the kinds in question we also do such things as assert of something that it exists in a certain mode, where to assert this is at least to assert (or anyhow imply) that subject-predicate statements the subject-expressions of which refer to the thing in question carry certain more or less specific sorts of implications and are verifiable and/or justifiable in certain more or less specific sorts of ways. Examples of sentences the uttering of which under certain circumstances would be to make assertions satisfying this description would be: "The ghost you 'saw' exists in your imagination," "The final battle of the Giants and the residents of Valhalla has only mythological existence," and "Her playmates, unlike Ramona's, really exist." Finally, both Shaffer and Baier add to this list of "existential speech acts," as they might be called, those in which by uttering sentences of the sorts in question it is asserted of a concept or an expression that it has a certain use or "logic." Sentences which can be uttered to perform acts of this type, it is suggested, would include: "Beasts of fable do not exist," "Particulars exist," and the like.

In suggesting, as I have, that these distinctions shed light on certain traditional questions concerning existence, I have particularly in mind the question—as it is usually framed—whether existence is a predicate, and the related question as to the truth and consequences of the view which is often expressed that the statement "God exists" is analytic. The first question is usually explicated as follows: when we say of a thing that it exists, is it the case that we are predicating a property of this thing or attributing a characteristic to it? Now, it has not been unusual for this question to be treated by simply noting, for example, that when a person says something like "Tame tigers exist," he is obviously not attributing a characteristic to a "subject" picked out by the expression "tame tigers"; rather he is talking about a concept—the

<sup>2</sup> There are of course many speech acts we perform by uttering sentences of the sort in question which are closely similar to those indicated: e.g., asserting of a *set* of individuals that they existed . . . , denying of an individual that it now exists, etc.

concept expressed by "tame tigers"—and saying of it that it has application to some thing or things. That this sort of treatment is too cavalier, however, becomes evident as soon as one is reminded of the distinctions noted above. It's simply not possible to construe all utterances of sentences like "The Boston Symphony Orchestra still exists" and "Her playmates really exist" as concept-application claims; hence the bearing of the points noted upon this traditional question.<sup>3</sup> The second question, upon which the distinctions mentioned bear, is treated with special clarity in Shaffer's paper. The central point is simply that once it is recognized that there are tautologous existence-claims—that is, that sometimes in uttering sentences like "A exists," etc., what we are doing (*inter alia*) is simply indicating that a necessary condition of a concept's having application to some thing is that the thing in question exist—it becomes clear both (a) that "God exists" can be employed to express a necessary truth, and also (b) that this fact in no way implies the cogency of any version of the Ontological Argument for God's existence.

Paradigm Case Arguments so-called take many different, though related, forms in the recent literature. Probably the most tempting, however, are not those which make explicit appeal to how we allegedly "can" or "do normally" or "ultimately must" learn and teach the meaning (use, sense) of some expression (and which hence involve a good deal of *a priori* psychology), but rather those which exemplify a form like the following:

1. It is a necessary (analytic, conceptual, logical) truth that "E" applies to any situation or items which display or possess features  $X_1, X_2, \dots, X_n$ , and which are in other respects normal—i.e., perfectly familiar, ordinary, everyday situations or items, free of the miraculous, untouched by monstrosity, etc.
2. There exist situations or items which . . . (as above).

Therefore,

3. There exist cases to which "E" truly applies.

(A particularly forceful example of an argument of this form is presented by Max Black in *Models and Metaphors* [Ithaca, N. Y., 1962], chap. 8). Now

the recent discussions, alluded to above, which seem to me to be among those which bear most significantly upon the question as to the value of arguments of this form are those contained in H. Putnam's essay "Dreaming and 'Depth Grammar'" in *Analytical Philosophy*, edited by R. J. Butler (Oxford, 1962) and P. Feyerabend's essay "Explanation, Reduction, and Empiricism" in *Minnesota Studies in the Philosophy of Science*, vol. 3 (Minneapolis, 1962).

Essential to the cogency of any argument of the form in question is the acceptability of the idea that no scientific discoveries about appropriate, perfectly familiar, normal (etc.) situations or items could reasonably affect the applicability of "E." (This point is made explicitly by Black in the course of presenting just such an argument: see Black, *Models and Metaphors*, pp. 158 ff.). Accepting this idea, however, is tantamount to holding that if, because of certain scientific findings or the development of certain theories, a person were to doubt or wonder whether "E" applies to a certain perfectly normal, everyday (etc.) item of the appropriate sort, then this person could properly be said to be changing the meaning of "E," using "E" with a new sense, or altering the concept expressed by "E." Now the point made in the former essay which is of most significance here is that one plausible criterion for saying that such and such an expression is *not* being used with a new or altered sense or meaning is the criterion which is satisfied when (a) a large number of speakers of the relevant language automatically come to speak in the allegedly deviant way, given the new scientific findings and/or theories, and (b) a large number of "hearers" of such talk "pass" it without feeling or detecting the slightest trace of linguistic aberration. This point is significant because if it is correct, then it follows that it is by no means clear that the view just noted—to which defenders of Paradigm Case Arguments of the form in question appear to be committed—is defensible. This is so because of the obvious difficulty in showing proleptically as regards some expression "E" that a criterion of the sort in question will never be satisfied.

<sup>3</sup> It should be noted in the present context that Alston's paper provides a considerably more subtle argument for denying that existence is a predicate than the sort indicated above, an argument which is by no means undone by reflection on the distinctions I have noted. However, since the argument turns—so far as I understand it—upon the claim that it is impossible to raise questions about an object of reference without believing (knowing, supposing, assuming, presupposing) that it has a certain more or less definite status as a mythological entity, or a spatio-temporal particular, or an item in somebody's dream, etc., it falls apart in view of the fact that, as Strawson has pointed out in *Individuals*, we can identify objects *relatively*, e.g., as simply the thing (person, event, etc.) so and so mentioned (described as generous beyond all comprehension, said to be more than ten tons, etc.).

The bearing of Feyerabend's essay on the question concerning the value of Paradigm Case Arguments of the form noted arises not out of any suggested difficulty involved in defending the first premiss of such arguments; rather it arises from a central claim in the essay which implies the irrelevance of such alleged analyticities to the existential conclusions of such arguments. The claim in question is that the concepts we employ in making statements about the way the world is are (or anyhow may be) "theory-laden," i.e., they are concepts which encapsulate (or anyhow may encapsulate) falsifiable theoretical elements (views, doctrines). And the claim, if true, drives a wedge between the allegedly analytic character of the first premiss of any Paradigm Case Argument of the form in question and the conclusion of such an argument just because it entails that a person might, under certain circumstances, be quite justified in accepting that such and such is analytic, but also justified in rejecting the very concept in which this truth is rooted in virtue of having good reason for rejecting the view(s) or doctrine(s) this concept encapsulates. The importance of Feyerabend's essay *vis-à-vis* the evaluation of Paradigm Case Arguments of the form in question lies not, it should perhaps be said, in the originality of this claim. Similar things have often, and often recently, been said. Its importance lies rather in the character and relative power of the argumentation adduced in support of the claim. I say "relative power," however, because a good deal would still have to be said before it would really be plausible, so far as I can see, to suggest that concepts like *action*, *pain*, and *anger* are as subject to alteration or elimination with the advance of knowledge as the concepts *phlogiston*, *impetus*, and *devil-possession*.

The consideration of the problem of existence as it arises especially in connection with scientific theory alluded to under (3) above is found in R. Harré's *Theories and Things: A Brief Study in Prescriptive Metaphysics* (London and New York, 1961). The main task Harré sets for himself in this book is that of adjudicating between "two systems of prescriptive metaphysics," one of which, called by Harré "positivism," involves such doctrines as that "the belief of people like C. T. R. Wilson . . . that the electron was discoverable . . . was an illusion" (p. 46), and the other of which, called

"the doctrine of ontological depth" (p. 2), is much less stringent than positivism as regards existence-claims. Harré argues in behalf of the latter "system," though in ways he admits it to be less than conclusive. One line of argument Harré advances runs in outline thus. On positivist principles it is impossible to find experimental grounds for ceasing to treat a model as "a work of the imagination" (p. 43) and beginning to view it as a "mechanism" which really exists. But there are at least two kinds of experiment which can legitimately be interpreted as providing such grounds. On the one hand, there are experiments which establish "family continuity" between things (e.g., tables and chairs) which have a reasonable claim to existence and certain others whose claim to existence has previously been precarious; and, on the other hand, there are experiments which yield support for a theory (and hence for the existence of the mechanism it posits) which explains the behavior, structure, etc., of such *prima facie* or anyhow well established existents. Hence positivism is unacceptable. To indicate exactly how Harré proceeds in order to make the premisses of this argument attractive would require more space than is available here. Suffice it to say that in the course of trying to do this he makes a number of provocative suggestions about the concept of existence with which we actually seem to operate, and that the discussion is helpfully concrete in virtue of numerous references to actual developments within the history of science.<sup>4</sup>

### III. CAUSALITY

Causal expressions such as "brought about," "in consequence of," "made," "effect," "was due to," "caused," "as a result of," etc., are employed in a large variety of contexts. They are employed, for example, to describe certain of the relations of extra-human natural events, various of the connections which often hold between human actions and subsequent states of affairs, the ways in which reasons and motives are upon occasion linked with conduct, relations which sometimes obtain between a man and the movements of his limbs, etc. In recent years a good deal of worthwhile work has been published concerning both the import of, and the principles governing,

<sup>4</sup> Good discussions of some of the issues surrounding the problem as to the status of "theoretical entities" may also be found in E. Nagel, *The Structure of Science* (New York & Burlingame, Harcourt, Brace & World, 1961), chap. 6, and Mary B. Hesse, *Forces and Fields* (London et al, Thomas Nelson and Sons, 1961), chap. 1 and *passim*.

our use of causal expressions in these different contexts. Especially worthy of note in this connection, it seems to me, are (1) the treatment given by H. L. A. Hart and A. M. Honoré in *Causation in the Law* (Oxford, 1959) of singular causal judgments in which some "contingency" is stated to be the effect (result, consequence) of some preceding "contingency" such as a human action or a natural occurrence; (2) the discussions in G. E. M. Anscombe's *Intention* (Ithaca, N.Y., 1957), S. Hampshire's *Thought and Action* (London, 1959), and A. I. Melden's *Free Action* (London, 1961) of the kind of causality—sometimes called "voluntary causality"<sup>6</sup>—expressible by statements of the form "A brought it about that p," where "A" is replaceable by the name of a person and "p" by a proposition such as "His arm went up"; (3) G. E. M. Anscombe's consideration in *Intention* of a certain class of causal judgments in which some mental event or something perceived by a person is asserted to be the "cause" or some feeling(s), thought(s), bodily movement(s), or action(s) of that person, and a typical member of which is "My having jumped that way was due to my having seen a face at the window"; and (4) Zeno Vendler's discussion of the logical differences between "effects, results, and consequences" in his paper under that title in *Analytical Philosophy*, edited by R. J. Butler (Oxford, 1962).

Among the main points which Hart and Honoré bring out concerning causal judgments like "The cause of the accident was the icy condition of the road" are these: first, that though such statements characteristically imply the truth of some causal generalization(s), the nature of such "principles" is different from what has traditionally been thought and their bearing upon singular causal judgments more complex than has traditionally been thought; and second, that the distinction which often requires to be drawn in causal contexts between the "cause" of some contingency and the "mere conditions" of the contingency's coming to pass has similarly been inadequately understood. In support of the first point, Hart and Honoré note, among other things, that the generalizations to which we appeal in backing our singular causal judgments are often not—as, e.g., Mill thought—statements of "invariable and unconditional sequence," i.e., statements of uniformities which hold "always" and under "all possible conditions." Rather they are statements of connections of a sort which are characteristically formulated in

very broad and very general terms and which are "like recipes in which we assert that doing one thing will 'under normal conditions' produce another" (Hart and Honoré, *Causation in the Law*, p. 29). To so characterize these general statements, however, is not, they insist, to imply that they are straightforwardly statistical in nature; for if they were straightforward statistical generalizations, they could not perform their role in justifying singular causal judgments. In addition, they note that typically singular causal judgments express connections which involve a succession of phases or stages each of which exemplifies a different causal generalization. Thus, "the statement that a slate falling from a house top caused the bruises of a passenger in an open car on whom it fell rests," they point out, "... on a *set* of both mechanical and physiological platitudes" (p. 43; italics mine).

By way of defending the second point noted above, Hart and Honoré argue with considerable plausibility that two contrasts are of central importance as regards the distinction in question—viz., the contrast between the abnormal and the normal, and the contrast between "contingencies" characterizable as free, deliberate human actions and those not so characterizable—and that these contrasts come into play in very subtle and complex ways in different sorts of contexts. Their argumentation is, of course, much too involved for summary here. But in the main it turns upon such considerations as, for example, (a) that though ordinarily the presence of oxygen is a "mere condition" of a fire's breaking out, in special circumstances where the presence of oxygen is abnormal—e.g., in a factory where special precautions are taken to exclude oxygen during some manufacturing process—this condition can be elevated to rank of cause, and (b) that the presence of large amounts of arsenic in a body is in many contexts displaced as the cause of death by a deliberate action, namely the poisoner's, an event which also characteristically constitutes the final link in the backward tracing of causal connections.

The discussions provided by Anscombe, Hampshire, and Melden which bear upon the question of the proper understanding of "voluntary causality" bring out, among other things, (1) innumerable difficulties in the way of understanding statements like "Brown made his stomach muscles contract" which is associated especially with the names of Hume and Mill, and according to which such

<sup>6</sup> By, e.g., P. Geach in "Ascriptivism," *Philosophical Review*, vol. 69 (1960), p. 224.

statements assert a connection of a certain kind between two events, the later being (typically) some bodily movement or anyhow some physical occurrence and the earlier being some mental event, frequently referred to as an "act of will," "a volition," "an act of intending," etc.; and (2) a large number of the significant logical features of the concept of an (intentional) action, which concept, it is suggested, is the crucial one to investigate by way of gaining a correct understanding of statements of so-called "voluntary causality." Among the difficulties which allegedly beset the Hume-Mill way of treating this kind of causality are the following. If in saying that Brown made his stomach muscles contract we are asserting that some volition (say) which occurred "in" Brown's mind had as a consequence the movement of certain muscles, then it would seem that either (a) this volition was something Brown brought to pass (made to happen), or (b) it just came to pass (happened) independently of any doing of Brown's. But if (a) is true, there seems to be no way of avoiding a very perplexing infinite regress; and if (b) is true, it would seem that Brown is self-contradictorily reduced to the status of a mere passive spectator of his own actions. Again, if the connection asserted to obtain in statements like the one about Brown is a causal connection of just the sort which holds between (say) Charles' decapitation and Charles' death—as the Hume-Mill approach insists—then we must be able to identify specific volitions in some way other than as just the volitions which regularly precede such and such occurrences. But this in fact we cannot do—hence the difficulty.

Perhaps the most important of the logical features of the concept of an action which the discussions under consideration illuminate have to do with what applying this concept in a particular case rules out and how this concept is logically connected to various others, and in particular to the notion of a reason for acting. As to the former, it is argued that correctly judging such and such a happening—such as Brown's making his muscles contract—to be an (intentional) action of Brown's rules out (*inter alia*) (a) the possibility that Brown was not aware that he was contracting his muscles,

(b) the possibility that he found out or came to know he was contracting his muscles by making some sort of observation(s) together with (perhaps) an inference from what he had observed, and also (c) the possibility that Brown can explain the happening in question by appealing to some prior happening, together with some causal generalization which links this latter happening with the former. As regards the connection of the concept of an action to the notion of a reason for acting, among the central points which emerge are these: (a) that the appropriate answer to the question "Why did you do that?" asked in connection with some action, though typically an answer which gives a reason for acting, will not invariably be of this type; (b) that such a typical answer, though it may not explicitly indicate a desire or want on the part of the agent, will at least generally do so implicitly; (c) that the reasons an agent gives for his actions must fit together over a reasonable stretch of time in order to count as genuine reasons; (d) that there is no way of showing that a universal connection holds between any action of a specifiable kind and certain reasons for or against performing it, such that these reasons have to be regarded as good ones; and (e) that when an agent is deciding what to do on a given occasion by considering the reasons for or against performing each of the possible actions open to him, it is (logically) impossible that he should at the same time ask himself the question: "Knowing myself as I do, what am I in fact likely to do?"<sup>6</sup>

As regards the class of judgments typified by "My having jumped that way was due to my having seen a face at the window," Miss Anscombe's main point is simply that the cause (or, as she also puts it, the causation) asserted to be operative by a judgment of this kind is "in the class of things known without observation." That is to say, it is maintained by Miss Anscombe, first, that we do not come to know that such a "mental cause" is responsible for some thought or reaction (say) on the basis of clues or signs discovered by observation or of finding that something is the case which shows that such and such is the cause of the thought or reaction in question; and second, that none the less there does exist the possibility of

\* A number of valuable recent discussions of points relating to many of those mentioned in regard to the concept of action are contained in Richard Taylor's "I Can," *Philosophical Review*, vol. 69 (1960); J. Austin's *Philosophical Papers*, edited by J. O. Urmson and G. J. Warnock (Oxford, Clarendon Press, 1961); Carl Ginet's "Can the Will Be Caused?," *Philosophical Review*, vol. 71 (1962); *Freedom and the Will*, edited by D. F. Pears (London, Macmillan & Co., Ltd., 1963); the symposium on Human Action published in *The Journal of Philosophy*, vol. 60 (1963); and Anthony Kenny's *Action, Emotion and Will* (London, Routledge and Kegan Paul, 1963).

being mistaken whenever one makes a causal claim of the sort in question. Accordingly, what is of special interest about this alleged class of singular causal judgments is the failure of its members to submit to treatment along traditional Humean lines, despite the evident similarity of such judgments to those which are most happily viewed along these lines.<sup>7</sup>

Among the central conclusions at which Vendler arrives in the paper cited above are, first, that effects are (or tend to be) events or processes, whereas results and consequences are (or tend to be) facts or fact-like entities; second, that the difference between the latter, i.e., results and consequences, arises mainly in connection with human actions, results being those states of affairs due to human action which are "considered in connection with the actual or possible intentions of the agent," consequences being those which are "considered in abstraction from such intention" (Butler [ed.], *Analytical Philosophy*, p. 11); and, third, that the answer to the question whether effects have causes is, paradoxically enough, "No." What is perhaps of most interest about this paper, however, is not the results at which Vendler arrives, but rather the way in which the arguments allegedly yielding these results involve notions—to most philosophers, I suspect, quite unfamiliar—endemic to contemporary linguistic theory—in particular, notions which have been central in the development of transformational grammar.<sup>8</sup>

#### IV. TIME

Among the problems which have traditionally been of concern to philosophers as regards the present topic, and in connection with certain ones on which considerable serious work has been done in recent years, are: first, those which center around the notion of "time's passage" or the "flow of time"; second, those—often closely

connected with the first—which concern the "nature" and/or "status" of the future and the past; and third, those—again intimately connected to these others—which are often referred to by the quite confusing phrases "the problem of the direction of time" and "the problem of time's arrow."

Probably the central problem concerning the notion of temporal passage has always been the problem, as it is often (though somewhat roughly) put, of whether the facts we (ostensibly) express in saying such things as "June 15, 1970, is still in the future," "World War II has been over for eight years now," and "Her thirty-seventh birthday is moving steadily nearer to the present, will one day be present, and then will recede into the past, becoming more and more remotely past as time goes by" are or are not capable of being fully expressed by statements which indicate merely that certain unchanging relations of temporal precedence and/or simultaneity obtain among certain events or dates. Now of special interest as regards this problem are several of the recent essays of A. N. Prior, who has been perhaps the most vigorous defender in recent years of the more or less unpopular view that "time's flow" cannot be adequately treated by one who restricts himself to timeless truths about the relations of before and after which obtain among the denizens (and periods) of time.<sup>9</sup>

In support of this view Prior has presented considerations of both "offensive" and "defensive" kinds. Perhaps the most suggestive instance of the former is his criticism in "The Formalities of Omniscience" (*Philosophy*, vol. 37 [1962]), and more fully in "Thank Goodness That's Over" (*Philosophy*, vol. 34 [1959]), of "eternalistic" translations of statements like "The 1960 examinations at Manchester are now over" (hereafter *S*) on the ground that to be (e.g.) pleased that *S* is not *eo ipso* to be pleased that *S'*, where *S'* is the state-

<sup>7</sup> Some acute objections to the notion of mental causality as Miss Anscombe develops it are contained in J. Teichmann, "Mental Cause and Effect," *Mind*, vol. 70 (1961) and J. Gaskin, "Mental Causes and Fear," *Mind*, vol. 71 (1962).

<sup>8</sup> An elementary account of this development is provided in H. A. Gleason, Jr., *An Introduction to Descriptive Linguistics* (2d ed. rev.; New York, Holt, Rinehart & Winston, 1961). For a much more sophisticated discussion of the character and virtues of the transformational approach, as well as the ideas it employs, see Noam Chomsky, *Syntactic Structures* (The Hague, Netherlands, Moulton & Co., 1957). Able comments on Vendler's paper by Sylvain Bromberger and William Dray, together with Vendler's reply, are also contained in Butler (ed.), *Analytical Philosophy* (Oxford, 1962).

<sup>9</sup> Though also of interest in this connection are the following: R. Gale, "Tensed Statements," *Philosophical Quarterly*, vol. 12 (1962); Richard Taylor, *Metaphysics* (Eaglewood Cliffs, N.J., Prentice-Hall, Inc., 1963), chap. 6; G. J. Whitrow, *The Natural Philosophy of Time* (London & Edinburgh, Thomas Nelson & Sons, Ltd., 1961); W. Sellars, "Time and the World Order," *Minnesota Studies in the Philosophy of Science*, vol. 3, edited by H. Feigl and G. Maxwell, (Minneapolis, University of Minnesota Press, 1962), sec. 3; and Milč Čapek, *The Philosophical Impact of Contemporary Physics* (Princeton, N.J., D. Van Nostrand Company, Inc., 1961), chap. XVII and *passim*.

ment: "The exams are [in the timeless sense 'are' has in '2 and 2 are 4'] earlier than July 1, 1963" (I write this on July 1, 1963), or, alternatively, the statement: "The exams are [again in the timeless sense] earlier than the date of this inscription, whatever it is."

Among the most interesting of the defensive considerations Prior adduces are (a) those designed to rebut the charge that the view he espouses involves commitment to an infinite sequence of meta-times and (b) those designed to deal with such ostensible difficulties as that the view he defends implies that persons can change even when they don't exist. The former charge has its source in the consideration that if, e.g., V-J Day is steadily receding into the more and more remote past, then it must be moving at some constant rate; and if it is moving at some constant rate, then it must be covering such and such a temporal distance during every temporal period of such and such length, since this is what it is to move at a constant rate. But if this is what is involved in the steady recession of V-J Day, it would seem that the passage of time of which this recession is an example requires a second, higher-order time in which this passage takes place. And if this second time passes, as it must, then a third time will also be required, and so on *ad infinitum*. Prior's rebuttal of this charge in "Time after Time" (*Mind*, vol. 67 [1958]) consists in pointing out that there are at least two different places in the above reasoning where we can legitimately refuse to go along. Thus we can, he implies, simply refuse to say that V-J Day's recession entails that V-J Day is moving into the more and more remote past at some rate. And also, he implies, we can simply refuse to say that something's moving into the past at a certain rate—if we decide to speak in this way—involves any further time above and beyond the one familiar time in which ordinary events take place, and this because we can, if we wish, specify the rate at which V-J Day recedes as: one day per day, or one year per year, etc.

The difficulty concerning the alterability of people who no longer exist arises (I think) in view of considerations like these: at one point in time it was a fact, according to the view Prior defends, that Queen Anne died three years ago; at a later point, however, this fact no longer held. Now it's

natural to think of this fact as being a fact about Queen Anne. But to say that a fact about a person holds at one time and fails to hold at a later time is tantamount to saying that a change has taken place in that person. Hence the temptation to think that the view in question implies that persons change even when they no longer exist. One of Prior's ways of circumventing this ostensible difficulty in "Changes in Events and Changes in Things" (*The Lindley Lecture* [Lawrence, Kansas, 1962]) is to insist that we can understand the perplexity-engendering facts in such a way that persons do not "enter into them" at all. Thus, he suggests, "The fact that Queen Anne died just three years ago no longer holds" can be construed without significant semantic loss or gain by some such general statement as: "It was the case that  $(\exists x)(x \text{ was called 'Anne,' reigned over England, etc. \& } x \text{ died three years ago})$  and it is not now the case that  $(\exists x)(x \text{ was called 'Anne,' reigned over England, etc. \& } x \text{ died three years ago})$ "—a statement which, allegedly, is much less easily viewed as a statement about a particular person.

Among the problems which—roughly speaking—have to do with the "nature" and/or "status" of the future and the past, and which have of late been treated in a relatively fresh way are (1) the problem as to what the alleged "openness" or "ambiguity" of the future consists in and (2) the problem as to whether past things and events are, as has often been maintained, only contingently related to "the contents" of the present.

The treatment of the first problem which I have in mind is that presented by B. Mayo in his long and subtle paper on "The Open Future" (*Mind*, vol. 71 [1962]). His suggestion is that what our sense of the future's "openness" reflects is essentially the fact—recently expressed also by G. Ryle and A. N. Prior<sup>10</sup>—that future-tensed statements, unlike present- and past-tensed statements, are ineluctably general, or more fully, the fact that it is impossible to assert a proposition expressed by a future-tensed subject-predicate sentence, the subject term of which is singular. In defense of his claim that there is the distinction noted, however, Mayo argues (a) that a necessary condition of the assertibility of a proposition of the kind just described is that "the subject term, which purports to name some particular thing, must actually

<sup>10</sup> See G. Ryle, *Dilemmas* (Cambridge, Cambridge University Press, 1954), chap. 2, and A. N. Prior, *Time and Modality* (Oxford, Clarendon Press, 1957), *passim*, and "Identifiable Individuals," *Review of Metaphysics*, vol. 13 (1960). A more careful expression of this alleged fact may also be found in J. Margolis, "Statements about the Past and the Future," *Philosophical Review*, vol. 72 (1963).

name some particular thing: it must be given a reference" (*ibid.*, p. 13) and (b) that this condition is never satisfied in the case of propositions of the kind in question; and this line of argument forces him, so far as I can see, into having to accept such absurdities as that, in the case of a statement like "This dirty building will be clean on Tuesday," "this dirty building" cannot function successfully as a referring expression. None the less, the paper is a rich and suggestive one, and a truth of some importance is, I think, at least adumbrated in it.

Equally interesting is N. Malcolm's recent essay on "Memory and the Past" (*Monist*, vol. 67 [1963]), in which are considered a number of different ideas that bear upon the second problem just indicated. They are, roughly expressed and in the order of their treatment: (a) the idea of the world's having come into being five minutes ago complete with a population that *remembers*, in Russell's words, "a wholly unreal past"; (b) the idea of the (or a?) world's springing into being complete with a population that seems to remember "a wholly unreal past"; (c) the idea of the (or a?) world's springing into being complete with a set of persons who, over a considerable period of time "exhibit normal memory-reactions in daily behavior and also gradually demonstrate a mastery of language, including the past tense," and who then "begin to express, in the language which they understand, apparent memories of 'a wholly unreal past'" (p. 261); and (d) the idea that our world (i.e., the planet Earth) and mankind sprang into existence five minutes ago. Of special interest in his discussion of these ideas is the argument that (b) fails to express a "logical possibility" on the ground that this idea involves the notion of an entire population of persons the vast majority of whose statements about the past are false, a notion which is incoherent since the general falsity of their statements about the past would deprive us of any basis for thinking they understand (or have) the past tense. Especially perplexing, however, is the argument that (c) is similarly defective on the ground that (i) if the apparent memories of the members of the population in question agreed with one another and with their "records" (Malcolm's quotes), then their apparent memories would be verified as true; and (ii) if this were so, it would not be intelligible to hold that the past they describe may not have existed. This argument is perplexing because what shows a memory to be veridical is, one

naturally supposes, its agreement with other memories and/or with certain records, not, as Malcolm says, its agreement with other *apparent* memories and/or with certain "records."<sup>11</sup>

Lastly, I should like to mention in connection with the so-called "problem of time's arrow" Adolf Grünbaum's many-faceted essay "The Nature of Time" in *Frontiers in Science and Philosophy* edited by Robert G. Colodny (Pittsburgh, 1962). This essay is, I think, particularly helpful owing to the way in which it both distinguishes, and yet brings out possible connections between, a number of different problems which have often been more or less muddled together in the literature concerned with this problem-area. Thus, among others, Grünbaum treats in this essay:

1. The problem of discovering some feature(s) of "physical reality" by reference to which it is possible to "explain" the putative facts which constitute "time's flow" or "passage";
2. The problem of discovering some feature(s) of "physical reality" by reference to which it is possible to "explain" the facts we express when we say truly that, for example, some event occurred *before* or *after* some other; and
3. The problem of discovering some explanation of the existence of those processes that we find empirically to be irreversible which appeals to acceptable physical theory alone.

Of particular interest is Grünbaum's suggestion that the solution he propounds to the third of these problems can also be appealed to in treating both the second and the first—the second because the fact that our "experiences" are ordered by such relations as *earlier than* is simply a reflection on the "psychological plane" of processes of the type accounted for (and hence required) by the principle set forth in solving problem 3; the first because, as Grünbaum puts it, "it is a matter of fact that the Now 'shifts' in conscious awareness to the extent that there is a *diversity* of the Now-contents. But since these diverse Now-contents are ordered with respect to the relation 'earlier than' no less than with respect to its converse 'later than,' it is a mere *tautology* to say that the Now shifts *from earlier to later*" (p. 154: *italics* his). Whether Grünbaum's solutions of these latter problems are capable of withstanding careful scrutiny of (a) the mental-physical dualism his language seems to presuppose and (b) the notion that "the Now shifts *from earlier to later*," is, I think, a real

<sup>11</sup> This point was suggested to me by my colleague F. A. Siegler.

question—though not one which can be profitably tackled here.<sup>12</sup>

### V. PERSONS

The contributions I want to mention under this heading all bear on the age-old discussion of the relations between psychological facts, or facts about mental phenomena, and physical facts, or facts concerning bodily phenomena. They fall, however, into four more or less distinct groups: the first consists of an argument designed to undercut the view that a person is a compound particular, one component being a physical thing or collection of physical things, and the other being a non-physical thing or a collection of non-physical things; the second consists of several quite original arguments in support of the now familiar—though much-disputed—thesis that certain mental and physical phenomena are “logically” or conceptually connected; the third consists of a number of points which have been made of late by way of meeting objections to the “Identity theory” of mind and brain, i.e., the view that mental items, such as pains, experiences of seeing after-images, and the like,<sup>13</sup> might turn out to be, as a matter of empirical fact or plausible scientific theory, identical with certain brain events or processes; the fourth consists of a recent attempt to “analyze” mental items like pains, tickles, itches, and other such “bodily sensations” in such a fashion as to make more promising a somewhat different route to metaphysical materialism than that taken by recent Identity theorists.

The argument in criticism of the theory that persons are analyzable into particulars of more fundamental kinds is presented in Strawson's *Individuals*. In outline it runs as follows. We can and do ascribe to ourselves not only characteristics of a kind frequently exhibited by things to which we wouldn't dream of ascribing consciousness, but also characteristics of a kind the possession of which entails the possession of consciousness on the part of their possessors. A necessary condition of our doing this, however, is that we should be “ready and able” to apply the appropriate predicates to things other than ourselves. That is, we must know

(have a clear idea) under what circumstances we could truly apply such predicates to other individuals. This is so in virtue of the very notion of a predicate, this notion being “correlative with that of a *range* of distinguishable individuals of which the predicate can be significantly, though not necessarily truly, affirmed” (p. 99, n. 1). But a necessary condition of our being thus prepared to ascribe such characteristics to others is that we be able to make identifying references to individuals of the right sort. If an individual is a particular, however, it is necessary in order to make an identifying reference to it that one know an individuating, empirical fact about that particular. Otherwise, one simply is not in position satisfactorily to answer the question “Which one do you mean?” But now it is in general possible to know such facts about particulars only provided that they either are locatable in both space and time (and hence are more or less straightforwardly physical in nature) or are uniquely related to, and hence identifiable by reference to, such spatio-temporal particulars. The latter possibility as regards the individuals in question is, however, ruled out in view (*inter alia*) of the impossibility of knowing that there exists one and only one particular of the right sort in relation to any specified physical object. Hence, a necessary condition of ascribing experiences, or states of consciousness, to oneself is that the entity to which one ascribes them be a spatio-temporal particular—and *a fortiori* the very same entity as that to which one ascribes such things as weight, position, and color when one ascribes these characteristics to oneself. This summary statement of Strawson's argument by no means reflects its full complexity. However, it does perhaps suggest something of the argument's ingenuity and originality.

Among the more original arguments designed to undercut the view that all psychological facts are logically or conceptually independent of facts about human bodies and their movements are a number of those presented in Sydney Shoemaker's recently published *Self-Knowledge and Self-Identity* (Ithaca, N.Y., 1963) and the following argument of Strawson's, presented again in *Individuals*. Not all of our ways of telling on bodily grounds that

<sup>12</sup> For quite different recent approaches to the problem-complex in question see Richard Schlegel, *Time and the Physical World* (East Lansing, Michigan, Michigan State University Press, 1961), chap. 2; Max Black, *Models and Metaphors* (Ithaca, N. Y., Cornell University Press, 1962), chap. 10, and D. S. Shwayder, “The Temporal Order,” *Philosophical Quarterly*, vol. 10 (1960).

<sup>13</sup> Exactly what mental entities are held to be at least in principle identifiable with brain events or processes varies somewhat from theorist to theorist.

such and such a person is now undergoing such and such an experience can be mere "signs" of the experience in question, Strawson says. For if they were, then presumably we should have had to discover their epistemic value by observing correlations in our own case between the "occurrence" of experiences of the sort in question and the holding of the appropriate bodily facts. But a necessary condition of discovering such correlations to obtain in our own case is our being able to attribute to ourselves the experiences in question. However, for reasons of a sort indicated above, we could do this, Strawson insists, only provided that we were prepared to ascribe such experiences to things other than ourselves. Hence, a necessary condition of its being true that all of our ascriptions of experience or states of consciousness to others are inductively based is that it be false that all are inductively based. And this establishes that it is false that all such ascriptions are inductively based.

One of the more inventive arguments presented by Shoemaker to the same conclusion runs in outline thus. If all psychophysical relationships were contingent, then it would be impossible to have any knowledge of the "minds" of others. This is so because at best one could come by such knowledge in virtue of having discovered certain correlations between psychological and bodily facts to hold in one's own case. But in fact such correlations could never be discovered. How, Shoemaker asks, could I obtain knowledge that the relevant bodily facts hold? I can't, for example, rely on my eyes if it is a contingent fact that "visual experiences 'correspond' to objective physical facts in the way they do, i.e., that in general one seems to see (has the experience of seeing) an object of a certain kind only when one's eyes are directed toward an object of that kind" (Shoemaker, *Self-Knowledge and Self-Identity*, p. 192); and similar points can be made in connection with the other possible replies. But, the argument continues, we do have knowledge of the "minds" of others. Consider the sentence *S*: "It is logically impossible for one person to know of another that he is in pain." Now, Shoemaker points out, if *S* expresses a statement, then it expresses a statement that implies that "pain" does not have "an established meaning" in English. For suppose "pain" does have such a meaning. Then it would be possible to know that another person understands its meaning, i.e., uses it correctly. But a necessary condition of this being possible is that

we can tell if another is in pain, since essential to establishing that another understands "pain" is seeing that he utters sentences like "I am in pain" with the intention of making true assertions only when he is in pain. Hence, there is an absurdity involved in anyone's uttering sentence *S* with the intention of making a true statement, from which it follows that we do know certain "psychological" facts about others. But if this is so, the argument concludes, then clearly it is false that all psychophysical relationships are contingent ones.

The more interesting and suggestive points which have been made of late by way of defending the "Identity Theory" include the following. First, there is the point made by U. T. Place (in "Is Consciousness a Brain Process?" (*British Journal of Psychology*, vol. 47 [1956]), and based upon consideration of various identifications made in the history of science, that it is *not* a necessary condition of two items being discovered empirically to be identical that their identity be determinable simply by more careful observation of one or both of them. Second, it has been suggested by J. J. C. Smart in "Sensations and Brain Processes" (*Philosophical Review*, vol. 68 [1959]), and argued at length by J. Shaffer in "Could Mental States Be Brain Processes?" (*Journal of Philosophy*, vol. 58 [1961]), that the mere fact that mental items such as thoughts and experiences of pain are not locatable in physical space—at least as things now stand semantically—is no serious bar to their being in fact identical with certain physical phenomena, because we can simply adopt certain spatial conventions to govern the words which express the appropriate mental concepts. Third, in reply to the objection sometimes raised that sentences like "The pain I now feel in my back is (identical with) the neural discharges occurring in such and such region of my brain" are as deviant as N. Chomsky's "Colorless green ideas sleep furiously," and hence not the sort of linguistic entity which could be uttered to make a true statement, it has been argued by H. Putnam (in "Mind and Machines" in *Dimensions of Mind*, edited by S. Hook [Washington Square, N.Y., 1960]) that sentences can become non-deviant with the growth of knowledge, and this without its being obviously true that any of their component words have undergone a change of meaning. Finally, it has been suggested by Smart (in the paper indicated above) that what we are really saying when we report that we are, for example, now having an experience of seeing a yellow after-

image, is that there is now taking place in us some internal process or other which is similar to the process which occurs when, for example, we see a yellow object of a similar shape—a suggestion designed to put down various difficulties which turn upon the alleged fact that the identification of such items as experiences of seeing after-images requires reference to one or more peculiarly mental features of such items.

The attempt to “analyze” bodily sensations like pains, aches, and itches to which I alluded above and to a consideration of which I now turn by way of concluding the present discussion is that made by D. M. Armstrong in *Bodily Sensations* (London, 1962). The view for which he argues is, roughly, that such a sensation is “really” a “bodily sense-impression” (together with, in some instances, a certain “attitude” and/or certain behavioral tendencies caused to exist by the impression in question), the notion of a bodily sense-impression being introduced by reference to statements of the form: “I seem to (or: It seems to me that I) perceive by bodily sense a disturbance of some (more or less indeterminate) sort going on in such and such a place in my body,” where the “seems”-locution is understood in its “phenomenological sense.” Among the considerations advanced in support of this more or less sophisticated sense-datum analysis of pains *et al* are these. The view accounts for the privacy of such sensations, i.e., for the fact that no one can “have” anyone else’s sensations; for the impossibility of their “existing” unfelt by anyone; for the logical impossibility of our being *radically* mistaken about their “existence” or their “locations”; and for the connection between the “place” a pain (for example) is felt to be and the place within the body of the disturbance which is normally spoken of as the “cause” of the pain in question. Many will, of course, feel that more needs to be said than Armstrong says about the alleged “phenomenological sense” of “seems”-locutions; others will, I think, balk as soon as they have applied to his analyses of specific sensations Moore’s open question test. None the less, *Bodily Sensations* is a worthwhile book on an important topic, and one can learn much from it.

## VI. IDENTITY

Among the items of special interest as regards the general topic of identity, mention may be made of: Vere Chappell’s careful and helpful treatment in

his paper “Sameness and Change” (*Philosophical Review*, vol. 69 [1960]) of some of the classical confusions arising out of a failure to understand the different “uses” of “same” and, in particular, the connection of these to the distinction between “referring” and “describing”; G. C. Nerlich’s paper “On Evidence for Identity” (*Australasian Journal of Philosophy*, vol. 37 [1959]) in which it is argued that similarity provides no basis whatever for a judgment that, for example, this is the same particular as the one which we saw yesterday, given spatio-temporal discontinuity of the former and the latter; B. A. O. Williams’ argument for, together with Anthony Quinton’s argument against, the thesis that bodily continuity is a (logically) necessary condition of personal identity—arguments which appear respectively in “Bodily Continuity and Personal Identity: A Reply” (*Analysis*, vol. 21 [1960]) and “The Soul” (*The Journal of Philosophy*, vol. 59 [1962]); and various of the points concerning identity made *passim* in Hampshire’s *Thought and Action*. However, I shall consider in some detail only a few of the more striking arguments on the present topic which are set forth in the above-mentioned books by Strawson and Shoemaker. In particular, I want to draw attention to (1) the intriguing and difficult chain of argumentation by which Strawson tries to establish the conclusion that qualitative identity or similarity counts as “logically adequate” evidence of particular-identity—at least under certain circumstances, and (2) a number of the arguments Shoemaker advances concerning the question as to the “criteria” of personal identity.

Strawson’s argument to the conclusion noted may be viewed as beginning with the reminder that we do make identifying references to particulars. Sometimes, Strawson points out, such references are effected by the use of demonstratives. This, however, is possible only when the particular in question is within (or very nearly within) the range of particulars sensibly present to the speaker at the time of the reference. When this condition is not met, it is necessary to rely ultimately upon the use of descriptions (names, he says, are “worthless without a backing of descriptions which can be produced on demand” [Strawson, *Individuals*, p. 20]). For such non-demonstrative reference to succeed, Strawson then claims, the speaker must be able to say which particular he means, and he can do this—at least in general—only provided he can give a description which uniquely relates the particular in question to some demonstratively

identifiable particular. This means, however, that in order to refer to the items we do, we must—in general—be able either to specify the spatio-temporal positions of these items with respect to our current spatio-temporal position, or to specify some unique relation(s) such items bear to items which are thus locatable. And this because only the system of spatial and temporal relations has the “peculiar comprehensiveness and pervasiveness” required for any such system “to serve as the framework within which we can organize our individuating thought about particulars” (p. 25). Thus a necessary condition, Strawson concludes, of our making the multifarious references we do is that we “operate with the scheme of a single, unified spatio-temporal system” (p. 31), i.e., a system every one of the (spatial) occupants of which is at any time spatially related to every other at every time. But now, the argument continues, a condition of our possession and use of this scheme is that we have criteria by reference to which we can reidentify items which endure through time. And, furthermore, any account of such criteria must, Strawson points out, include among them mere similarity or qualitative identity. This is so because if we were never able to ascribe particular-identity in cases where we had only qualitative identity to go on, then we should be unable to avoid operating with the idea “of a new, a different, spatial system for each new continuous stretch of observation” (p. 35). The latter idea is not, however, the idea we must employ: “It is of the essence of the matter that we use the same framework on different occasions” (p. 32). *A fortiori*, similarity constitutes, under certain circumstances, a logically adequate criterion of particular-identity. Once again it has to be said that the foregoing is but a sketch which brings out certain strands of Strawson’s tightly woven and complex argument, a sketch which none the less comes closer, I hope, to portraiture than caricature.

One of the arguments concerning personal identity which Shoemaker advances and which is of particular interest—especially in relation to most traditional discussions of the epistemological problems concerning memory—is the following argument in support of the view that memory is a criterion of personal identity, i.e., that a person’s confident and sincere claim to remember having done so and so (e.g.) is non-inductive evidence that he in fact did so and so, and *a fortiori* non-inductive evidence of his identity with the person who did so and so. If memory, the argument

begins, were not a criterion of personal identity in the sense indicated, then our acceptance of a person’s confident and sincere claim to remember having performed a certain action as evidence that he in fact did do the thing in question would be inductively (empirically) based, i.e., based (presumably) upon our having discovered empirically that confident and sincere memory-claims are generally true. But this is tantamount to saying that memory’s not being a criterion of personal identity entails that it is only a contingent fact—a fact which could be otherwise—that confident and sincere memory-claims are generally true. But if it were only a contingent fact that confident and sincere memory-claims are generally true, the argument continues, then (among other things) it would follow that there would be no way of telling whether a person understands “remember” (and its cognates), since essential to determining whether such understanding is present is finding out whether “under optimum conditions the confident claims he [the person in question] makes by the use of the words [words like ‘remember’] are generally true” (Shoemaker, *Self-Knowledge and Self-Identity*, p. 231). But if we could never tell whether a person understands “remember” (and its cognates), then obviously we could never discover that the contingent fact in question holds, i.e., we could never discover that, as a matter of contingent fact, confident and sincere memory-claims are generally true. But this, the argument concludes, is a logically absurd supposition—presumably because its truth would mean, among other things, that we could never be in position, as we obviously are, to accept a person’s confident and sincere memory-claims to have done a certain thing as evidence that he did that thing. Hence memory must be a criterion of personal identity.

In addition to holding that memory is a criterion of personal identity on grounds such as the above, Shoemaker also maintains that bodily identity is such a criterion, too. In support of this view he produces a number of very original arguments, among which are the following. (1) Suppose, he says, bodily identity were not criterial (i.e., non-inductive) evidence of personal identity. Then, presumably, some kind of “psychological facts” must be. But in order to appeal to any psychological fact or facts as grounds for an assertion of personal identity, it would seem that we must be able to establish such identity on quite other grounds. For consider the following: Requisite to

finding out that any psychological fact holds regarding some person (that, e.g., he has such and such character traits, interests, knowledge) it is necessary to observe the person over a period of time. But this means we must be in a position to determine that the one observed at  $t_0$  is the same person as the one observed at  $t_1$  independently of psychological considerations. Again, our knowledge that certain psychological facts hold in a particular case must often be grounded on what the person in question says. But in order to know that a person has said such and such, it is essential that we be in a position to know (a) that all of the words uttered have been uttered by the same person, and (b) that this person understands the words in question. And it is difficult to see, Shoemaker says, how we could come to know either of these things unless we could rely on non-psychological facts to establish identity. But there is nothing except bodily identity that we could use as a non-psychological ground of personal identity, and hence bodily identity must be a criterion of personal identity.

(2) A less tempting, but equally original, argument runs roughly thus. Again suppose that bodily identity were not a criterion of personal identity. On this supposition the identity of a person would be logically independent of the identity of any particular body. If this were so, however, then there would seem to be no reason in logic why persons should not be constantly changing place in a discontinuous way. But in fact the notion that persons should behave thus does not express a logical possibility. For suppose we did so behave. Then it would be impossible for us to reidentify objects, to know that this thing is the same (numerically) as that previously encountered thing. This is so because if we were constantly changing place discontinuously, we could never be sure (i.e., directly observe) that the criterion of spatio-temporal continuity is ever satisfied by any given object from moment to moment. But if it were impossible for us to reidentify objects, then it would be impossible for us to use any object as a point of reference by relation to which the positions of other items (including ourselves) could be specified, and *a fortiori* it would be impossible for us ever to find out that we were constantly changing place discontinuously. But if nothing could show that this notion is true, it does not express a logical possibility. Hence the original supposition must be false, and bodily identity must be a criterion of personal identity.

## VII. God

There exist in the recent literature quite a number of books and essays which might usefully be mentioned under this last heading. There is, for example, the quite unusual "ontological argument" recently advanced by Frederick B. Fitch in "The Perfection of Perfection" (*Monist*, vol. 47 [1963]), which putatively shows that if it is possible that something is perfect, that "perfection is a possible attribute" (p. 468), then this attribute is non-empty, i.e., something is perfect, namely, the attribute itself which, it turns out, "cannot be merely an attribute" (p. 469), since being perfect it "must be whatever also is required of it in order to be perfect" (*ibid.*). There is the equally unusual "argument from design" in Richard Taylor's *Metaphysics* (Englewood Cliffs, N.J., 1963)—an argument which turns upon the thesis that, as Taylor puts it, "it would be irrational for one to say both that his sensory and cognitive faculties had a natural, non-purposeful origin and also that they reveal some truth with respect to something other than themselves, something that is not merely inferred from them" (pp. 100–101; italics his). There are the claims made by a number of philosophers of quite different orientations that there is a promising but so far largely untravelled path to religious truth which somehow runs between natural and revealed theology. Here one thinks, for example, of Ninian Smart's attempt to articulate and defend the notion of "religious reasons," i.e., considerations which, though necessarily indecisive (religion being what it is), can none the less lend rational support to one religious position rather than another (see, e.g., *Reasons and Faiths* [New York, 1959], *A Dialogue of Religions* [London, 1960], and, most recently, "Revelation, Reason and Religions," in *Prospect for Metaphysics*, edited by Ian Ramsey [London, 1961]). John Wild's recently propounded "Existential Argument for Transcendence" (in *Existence and the World of Freedom* [Englewood Cliffs, N. J., 1963], chap. 11) also comes to mind in this connection inasmuch as it is an argument which, he claims, avoids "the dogmatism of classical, natural theology" in virtue of not being an argument "in the traditional logical sense of either induction or deduction" (p. 203). And, of course, there is J. N. Findlay's very rich restatement of the central points of his famous paper "Can God's Existence Be Disproved?" in the ninth chapter of his recently published and extremely suggestive *Values and*

*Intentions* (London, 1961), not to mention the 27 avowedly question-begging "proofs of God" set forth in the fourth chapter of part I of Paul Weiss's massive and highly imaginative metaphysical treatise *Modes of Being* (Carbondale, Ill., 1958). However, perhaps the most notorious and certainly one of the most challenging contributions which has been made of late on the topic at hand is the line of reasoning in support of God's existence presented by Norman Malcolm in his essay "Anselm's Ontological Arguments" (*Philosophical Review*, vol. 69 [1960]).<sup>14</sup>

One way of setting out what seem to be the main steps of this argument—which has not, I think, always been adequately understood by commentators—is the following. The concept of God, as this concept is expressed by the most common use of the word "God," is the concept of a being "than which no greater can be conceived," i.e., of a supremely perfect being. But the concept of a being which is supremely perfect is the concept of a being which is, among other things, "absolutely unlimited." Whatever is conceived in these ways, however, is necessarily such that it can't be brought into existence by anything, it can't be caused to go out of existence by anything, it can't be prevented from existing by anything, it can't be kept in existence by anything, it can't just happen to come into existence or go out of existence, it can't just happen to exist or just happen not to exist, and it doesn't exist in or endure through time. Thus suppose a supremely perfect being could be prevented from existing by something. Then obviously the latter thing would constitute a limitation on the being in question. The possibility that there could be such a limitation, however, is ruled out by the fact that a supremely perfect being would necessarily be absolutely unlimited. Again, suppose such a supremely perfect and absolutely unlimited being were temporal, i.e., possessed duration in time. Then it would necessarily be capable of ceasing to exist, and also it could (logically) be caused by something to cease to exist. But anything with these features would be a limited being, which contradicts the original supposition. And so on. But now, the argument continues, anything conceived in such a way as

necessarily to have features like those just indicated logically cannot fail to have the additional "property" (hereafter *P*) that *either* it exists and its non-existence is logically impossible, *or* it fails to exist and its existence is logically impossible. For suppose something conceived in the way indicated lacked *P*. Then it would be so conceived that if it were to exist, its not having existed would none the less be logically possible. But if this last were so it would be so conceived that if it were to exist, it would depend upon other things both for coming into existence and for continuing in existence, and it would fail to be an eternal, i.e., non-temporal, being. But these possibilities are logically ruled out *ex hypothesi*. Hence any being conceived in the way indicated must possess *P*. Again, something conceived as possessing all of the various features noted above would *eo ipso* be thought of as being such that if it failed to exist, it couldn't come into existence, since it couldn't be caused to come into existence and it couldn't just happen to come into existence. Similarly, it would *eo ipso* be thought of as being such that if it existed, it couldn't cease to exist, since it couldn't be caused to cease to exist, nor could it just happen to go out of existence. But to conceive of something in these ways is to conceive it as possessing *P*. The concept of God expressed in the most common use of the word, then, is the concept of something which possesses the property *P*. But this is to say the concept is such that if God's existence is (logically) possible, then God exists. Now, the argument concludes, it is not possible to prove that the concept of God is not self-contradictory. However, in view of the fact that the concept has "a place in the thinking and the lives of human beings" (p. 60), there is clearly no more of a presumption that the concept is self-contradictory than there is that such familiar notions as that of seeing a material thing are self-contradictory. Hence the rationality of holding that God does in fact exist.

What helps to make this line of argument of special interest, I think, is the fact that it avoids at least two of the most standard criticisms of ontological arguments of more familiar types. One of these is the objection that such arguments establish at best only the hypothetical conclusion

<sup>14</sup> A quite similar argument is elaborately defended by Charles Hartshorne in the second chapter of *The Logic of Perfection and Other Essays in Neoclassical Metaphysics* (LaSalle, Illinois, Open Court Publishing Company, 1962). In this connection it is worth pointing out (a) that as regards at least the crucial question as to whether the concept of God is one which is logically capable of having application to anything Hartshorne's discussion is clearly superior to Malcolm's; and (b) that the distinction Malcolm makes between the two different lines of reasoning in support of belief in God in the *Proslogion* has been stressed by Hartshorne for many years.

that *if* anything satisfies the concept of God, then that thing exists. Now it's clear, I think, from the foregoing exposition of the main steps of Malcolm's argument that it is not open to an objection of this sort. This is so because the later steps of the argument involve a straightforward application of the rule of *modus ponens* to the hypothetical proposition "If God's existence is logically possible, then he exists" and the categorical proposition "God's existence is logically possible." A second familiar objection which Malcolm's argument circumvents is the objection that such arguments assume the allegedly questionable doctrine that existence is a predicate. Some, indeed, have actually levelled this criticism at Malcolm's

argument—doubtless owing partly to certain confusions in Malcolm's presentation. However, this objection also is readily seen to be inapplicable once it is recognized that the crucial existential conclusion is alleged to follow from the fact that a certain concept is logically capable of having application, in conjunction with a set of premisses which entail that it is an "internal property" of this concept that if it is not self-contradictory (or in some other way logically absurd), then it must have application. The failure of these objections to apply to Malcolm's argument by no means entails, of course, that the argument is free of defects. Why, in my judgment, the argument fails, I have indicated at length elsewhere.<sup>16</sup>

*University of Chicago*

<sup>16</sup> See R. Coburn, "Professor Malcolm on God," *Australasian Journal of Philosophy*, vol. 41 (1963). Other papers containing discussions of or comments on Malcolm's paper are the following: R. E. Allen, "The Ontological Argument," *Philosophical Review*, vol. 70 (1961); R. Abelson, "Not Necessarily," *ibid.*; T. Penelhum, "On The Second Ontological Argument," *ibid.*; A. Plantinga, "A Valid Ontological Argument?," *ibid.*; P. Henle, "Uses of the Ontological Argument," *ibid.*; G. B. Matthews, "On Conceivability in Anselm and Malcolm," *ibid.*; J. Berg, "An Examination of the Ontological Proof," *Theoria*, vol. 27 (1961); T. Brown, "Professor Malcolm on 'Anselm's Ontological Arguments'," *Analysis*, vol. 22 (1961); W. J. Huggett, "The Nonexistence of Ontological Arguments," *Philosophical Review*, vol. 71 (1962); J. Shaffer, "Existence, Predication, and the Ontological Argument," *Mind*, vol. 71 (1962); J. Yolton, "Professor Malcolm on St. Anselm, Belief and Existence," *Philosophy*, vol. 36 (1961); and K. Baier, "Existence," *Proceedings of the Aristotelian Society*, vol. 61 (1960-61).

## IV. BELIEF AND DISPOSITION

ISAAC LEVI AND SIDNEY MORGENBESSER

AS usually presented the thesis that a belief is a disposition to act<sup>1</sup> has two virtues and three defects. It is a clear improvement over the Humean thesis that a belief is a lively idea, and it appears to be of considerable relevance for psychological doctrine and theories of rational inference.<sup>2</sup> Unfortunately, there is the debit side. "Disposition to" is not always clearly explicated, nor is the philosophical problem which the thesis is intended to solve always explicit. And when it is, concomitant philosophical gain seems slight. But despite these ills the thesis calls not for abandonment but for reinterpretation which we shall attempt to supply after we document our complaints and indicate what we take to be legitimate sources of puzzlement and the ways of their removal.

### I

Analyses of beliefs as dispositions have sometimes been invoked in attempts to determine what significant relation, if any, obtains between belief-predicates taken in an occurrent and in a non-occurrent sense.<sup>3</sup> Jones may be said to believe that the Empire State Building is in New York City throughout his adult life and still be said in another sense to have believed so only on a few occasions during that period. We clearly cannot say that Jones' belief taken in the latter occurrent

sense is related to his belief understood in the former non-occurrent sense in the way in which becoming blue is related to being blue. Instead, so we are told, occurrent beliefs are related to non-occurrent beliefs in the way dissolving is related to solubility.

Whatever the merits of this proposal (our remarks later in the paper have some indirect bearing on this matter), we still do not know why we should attempt to show that beliefs (in the non-occurrent sense) are dispositions to overt action or behavior; for the proposal construes non-occurrent beliefs as dispositions to occurrent beliefs, not dispositions to action. The motive for attempting to treat beliefs as dispositions to action or behavior eludes us.<sup>4</sup>

But perhaps only temporarily; the problem of intentionality has provided others and perhaps it can provide us with a motive for attempting to analyze belief as a disposition to action. Consider, for example, R. B. Braithwaite's advocacy of the view that "*A* sincerely believes that *p*" is equivalent to the conjunction of "*A* entertains *p*" and "*A* is disposed to act in a manner appropriate to the truth of *p*."<sup>5</sup> In this way, Braithwaite hoped to reduce questions about intentionality of beliefs to questions about intentionality of entertainment or understanding of a proposition. In "Belief and Action," he conjectured that Wittgenstein's

<sup>1</sup> We mean to consider both dispositions to action and dispositions to behavior but shall not attempt, except in passing, to offer a way to distinguish them. For our purpose, a precise distinction is not essential, although we shall often follow the traditional but nebulous view that behavior consists only of bodily motions whereas action is explainable in terms of the goal directed character of such motions. This distinction may be allowed as a first approximation sufficient for the purposes at hand.

<sup>2</sup> Thus, if the "conclusion" of an inference is the acquisition of a certain disposition to act, a theory of rational inference is an organon designed for the pursuit of practical wisdom. Rational belief becomes a function of the factors controlling rational action. Whether this implication holds for all conceivable analyses of belief as disposition to action is debatable. It does seem to hold, however, for the typical views currently held.

<sup>3</sup> The distinction between occurrent and non-occurrent terms is not always drawn in a uniform way. It may contrast between: (a) terms applying to entities localizable in time (or space and time) and those which do not so apply; (b) terms which apply to entities in time (or space and time) for only brief periods and those which do not; (c) terms which describe changes in the conditions of objects to which they apply and those which do not ("becomes red" vs. "is red"); (d) terms which describe evanescent changes and those which do not ("flashes" vs. "beams"). Since disposition terms are usually occurrent in sense (a) and may be so in sense (b) (an iron bar may be magnetic for a brief period) and since most authors treat disposition terms as non-occurrent, we shall usually take the distinction in sense (c) or (d) without distinguishing between them.

<sup>4</sup> None the less, we may conclude that analyses of beliefs as dispositions to action are concerned with belief in the non-occurrent sense. We shall, therefore, restrict our discussion to that sense.

<sup>5</sup> R. B. Braithwaite, "The Nature of Believing," *Proceedings of the Aristotelian Society*, vol. 33 (1932-33), pp. 129-146, and "Belief and Action," *Proceedings of the Aristotelian Society, Supplementary Volume 20* (1946), pp. 1-19.

suggestion that understanding be treated dispositionally could eliminate reference to intentionality in analyses of belief altogether.<sup>6</sup>

Although intentionality presents interesting questions, doubt can be raised as to the need for analyzing beliefs as dispositions to action if the aim is to explicate belief-predicates without appeal to intentionality. Following Israel Sheffler,<sup>7</sup> we define "*A* believes that *p*" to mean the same as "There is a *y* such that *A* believes-true *y* and *y* is a that-*p*" where "*y*" ranges over inscriptions, that-clauses are predicates applying to inscriptions which are rephrasals of the sentential content of such clauses, and "believes-true" is a two place predicate whose domain is the set of persons and counterdomain the set of inscriptions. This procedure eliminates variables ranging over intentional objects without appeal to dispositionality.

Together with Sheffler, we acknowledge the operational criteria for applying the term "believes-true" to be no clearer than those for "believes."<sup>8</sup> However, the problem of clarifying conditions of application of belief-predicates or of reconstructing the relations between believing and overt behavior can be distinguished from the problem of eliminating the need to refer to objects of intention. *A fortiori* analyses of beliefs as dispositions to action are unnecessary to successful treatment of the problem of intentionality.

However, even if we had not found Sheffler's arguments convincing, we would still balk at any view such as Braithwaite's which holds that "*A* believes that *p*" entails "*A* is disposed to act in a manner appropriate to the truth of *p*." Our reasons follow a further examination of Braithwaite's approach.<sup>9</sup>

More fully but still roughly speaking, Braithwaite's analysis of the disposition attribution

requires construing it as a congeries of conditionals of the following form:

Whenever *A* has springs of action *S* and at the same time is subject to external conditions *E*, then he acts in a manner which he judges would be optimal for satisfying *S* if *p* were true.<sup>10</sup>

Ignoring the circularities implicit in this analysis (since judgments of optimality will involve beliefs), we still face other problems which can perhaps be most clearly indicated by trying to specify what beliefs (in Braithwaite's sense) a rational man could have.<sup>11</sup> Given certain requirements for rational action which Braithwaite apparently accepts, it can be shown that no rational agent can believe any propositions save those propositions which are necessary. In other words, a rational man is committed to rationalism.

For remember that if an agent has a goal determined by his springs of action *S* (let us ignore how goals are determined by springs of action), whether he will act on a proposition *p* depends, if he is rational, on how well the evidence available to him supports *p*. Moreover, the degree of evidential support for *p* which he will require before acting on *p* will change with his springs of action. To use an example discussed by R. C. Jeffrey, we may be prepared to act as if a given vaccine is nonvirulent in order to inoculate monkeys and be warranted in doing so without being warranted in inoculating children with the vaccine.<sup>12</sup> We clearly cannot analyze "*A* believes that *p*" as entailing a disposition to act on *p* relative to some specific set of springs of action without admitting that "Jones believes the vaccine to be nontoxic" means something different if Jones has in mind the inoculation of children or monkeys.

<sup>6</sup> "Belief and Action," p. 7. Presumably dispositions to act are dispositions to goal directed behavior in a sense akin either to Bigelow-Rosenblueth-Wiener or Braithwaite-Nagel-Sommerhof which allegedly does not introduce intentionality through the back door.

<sup>7</sup> I. Sheffler, "Thought on Teleology," *The British Journal for the Philosophy of Science*, vol. 9 (1959), pp. 280-281.

<sup>8</sup> *Ibid.*, p. 283.

<sup>9</sup> The difficulties raised here concern the implications of Braithwaite's analysis for theories of rational belief and inference (see footnote 2). Consequently, the thrust of our comments is independent of the attempt to eliminate intentionality and should be felt by analyses similar to Braithwaite's regardless of the purposes to which they are put.

<sup>10</sup> Braithwaite, *op. cit.*, p. 10-11. The springs of action are "the intentions, desires, wants, motives, instinctive needs, drives of the believer at the time when an external occasion for the appropriate action arises." (*Ibid.*, p. 9). Braithwaite takes the springs of action to be internal states and events which are causes or part-causes of human behavior. He, therefore, stands in opposition to Ryle who uses dispositional analyses of belief and motive predicates as a means of avoiding reference to internal states and occurrences. We shall argue later that Ryle's procedure leads to circularities which Braithwaite's analysis avoids.

<sup>11</sup> Testing the thesis in this manner is obviously not unfair, though Braithwaite restricts his analysis to the notion of sincere belief, leaving open the question of the rationality of the belief.

<sup>12</sup> R. C. Jeffrey, "Valuation and Acceptance of Scientific Hypotheses," *Philosophy of Science*, vol. 23 (1956), p. 242. Our analysis differs from Jeffrey's.

To avoid this counterintuitive result, Braithwaite requires that a sincere believer that  $p$  have disposition to act on  $p$  *no matter what his springs of action may be*. Thus, one could not sincerely believe in the nonvirulency of the vaccine unless he would use it to inoculate children as well as monkeys. Note, however, that for any contingent proposition  $p$  on which action can be taken,<sup>13</sup> there is at least one objective relative to which a non-suicidal, rational agent would refuse to act as if  $p$  were true. Consider, for example, the following gamble on the truth of  $p$ : If the agent bets on  $p$  and  $p$  is true, he wins some paltry prize and if  $p$  is false, he forfeits his life. However, should he bet on not- $p$ , he stands to win or lose some minor stake. A rational agent would bet on the falsity of  $p$  as long as the evidence did not confirm  $p$  to the maximum degree. Hence, on Braithwaite's account, the agent could not rationally and sincerely believe that  $p$  where  $p$  is contingent.

This is patently an undesirable consequence and counts heavily against Braithwaite's and kindred theses. But still many would claim that Braithwaite is on the right road and that somehow or other a correct analysis of belief as a disposition to act must be at hand in order to solve a variety of philosophical problems about concept-formation, theory construction, and explanation. These claims are important and complex and deserve full and independent treatment.

## II

In *Concept of Mind*, Ryle defends the thesis that motives (including beliefs) are dispositions to act by accusing rival views of implying that imputation of motives would be "incapable of any direct testing and no reasonable person would put any reliance in any such imputation. It would be like water-divining in places where well-sinking was forbidden."<sup>14</sup> Here Ryle is displaying interest in the concept-formational problem of specifying conditions of applicability of belief-predicates (more generally, motive-predicates) in some privileged "observational language" capable of describing publicly observable behavior. Both Descartes and Ryle would agree that belief-predicates are not publicly observational. The

question is: What conditions for applying belief-predicates can be given in terms of predicates characterizing publicly observable behavior?

But Ryle is not motivated only to answer this question. For concern with the problem of concept-formation is often expressed by him in conjunction with interest in a quite different question. Examine, for instance, the following passage:

There is, however, a special point in drawing attention to the fact that many of the cardinal concepts in terms of which we describe specifically human behavior are dispositional concepts, since the vogue of the paramechanical legend has led many people to ignore the ways in which these concepts actually behave and to construe them instead as items in the description of occult causes and effects. Sentences embodying these dispositional words have been interpreted as being categorical reports of particular but unwitnessable matters of fact instead of being testable, open hypothetical and what I shall call "semi-hypothetical" statements. The old error of treating the term "Force" as denoting an occult force-exerting agency has been given up in the physical sciences, but its relatives are perhaps only moribund in biology.<sup>15</sup>

Here it seems that Ryle expresses opposition to the paramechanical legend not only because it countenances "unwitnessable matters of fact" but because it makes references to "occult causes and effects." And not here alone, for much of Ryle's discussion is devoted to showing that motives and beliefs are not causes of action but play a different role in the explanation of human behavior—a role which, he alleges, can be illuminated by treating them dispositionally. In other words, Ryle's objectives in analyzing beliefs and motives as dispositions to action are twofold: to provide a way of specifying conditions of applicability of belief and motive predicates in terms of a language describing overt behavior and to clarify the role of such terms in explanation of such behavior. Ryle's two motives suggest that the claim that beliefs are dispositions to action is to be understood in two distinct senses, but Ryle himself seems to understand the notion of a disposition in a more or less univocal way. The attribution of a disposition for him is tantamount to a hypothetical assertion or conjunction of such assertions, and in this respect, his account of

<sup>13</sup> We should like to be able to say "any contingent proposition" but are aware of the difficulty of interpreting the concept of a bet on the truth value of some types of contingent propositions—most notably universal generalizations. However, we feel that Braithwaite's analysis falters on such propositions anyway.

<sup>14</sup> G. Ryle, *The Concept of Mind* (New York, 1949), p. 90.

<sup>15</sup> *Ibid.*, p. 117.

dispositions resembles Carnap's.<sup>16</sup> But we must confess that we have oversimplified. For the truth is that Ryle is actually bothered by two concept-formational problems, one of which stems from the non-observationality of belief-predicates and the other which is occasioned by other questions also considered by Nelson Goodman in his treatment of disposition-predicates. Patently, it is best not to treat all these issues concurrently, and hence discussion of Goodman's problem will be deferred until we have considered Ryle's claim that requests for conditions of applicability for belief-predicates, which are occasioned by their non-observational status, can be gratified by taking beliefs to be dispositions to action.

But what will be gained by such a maneuver? Certainly there is no inherent clarity about dispositional terms. The conditions of applicability of terms that almost everyone would have pre-systematically acknowledged to be dispositional such as "soluble," "brittle," "magnetizable," "possessing dormitive powers," etc., are as problematic as those for belief predicates. Moreover, Carnap's and Ryle's solution to the concept formational problem for dispositional predicates requires that the "stimulus" and "response" conditions in the hypotheticals into which disposition predicates are unpacked be specified in some privileged observation language. But a clearly dispositional predicate like "reflector of electromagnetic waves" does not seem to meet this requirement. Hence, the Carnap-Ryle account of dispositions rules out as non-dispositional predicates which presystematically appear clearly to be dispositional.<sup>17</sup>

Finally, terms like "has temperature *k*" which are not presystematically dispositional (at least not clearly so) are dispositional according to the Carnap-Ryle account in virtue of the fact that observable stimuli and responses are specifiable

for applying such terms in the same sense in which they are specifiable for "soluble."

These difficulties can easily be undercut by following the procedure apparently adopted by Carnap: namely, to construe the notion of a disposition predicate in a technical sense relevant to the problems of concept formation. According to this approach, all non-observational terms may be partitioned into two categories: theoretical and dispositional terms.<sup>18</sup> This distinction is drawn in terms of differences in the way in which non-observational predicates are interpreted in a privileged observational language.

This procedure avoids the reproach that justice is not being done to presystematic usage; but those who analyze beliefs as dispositions (in the technical sense) to action are faced with the following dilemma: Either they recognize theoretical (i.e., non-dispositional and non-observational) terms as scientifically legitimate or they insist that all non-observational terms be dispositional. If they opt for the first alternative, there seems to be no crucial concept formational reason why belief-predicates have to be treated dispositionally. Indeed, as Carnap has pointed out in a similar context,<sup>19</sup> strong insistence on such a procedure unnecessarily reduces the possibilities for theory construction in psychology. It might even be doing an injustice to our common sense theories; Whether a term is dispositional in the technical sense is determined in part by the theories in which it occurs. A commitment to treat a term as dispositional imposes too strong requirements on the theories in which it appears and might very well misrepresent the crude theories of common sense.

If we opt for the second alternative and insist (as Carnap seems to in *Testability and Meaning*) that all non-observational terms are dispositional, then the claim that belief predicates, which are

<sup>16</sup> Carnap mentions two possible rules for introducing dispositional predicates in terms of hypotheticals involving "stimulus" and "response" conditions. ("Stimulus" and "response" are to be understood to be antecedent and consequent conditions generally and not in a more narrowly psychological sense.) One is via reduction sentences; the other is via explicit definition in terms of hypotheticals involving causal modalities. R. Carnap, "The Methodological Character of Theoretical Concepts," *Minnesota Studies in the Philosophy of Science*, vol. I (Minneapolis, 1956), pp. 63-64. Ryle's account follows, roughly speaking, the second alternative.

<sup>17</sup> Carnap has virtually admitted that his account of disposition terms violates presystematic usage by baptizing terms like "reflector of electromagnetic waves," "theoretical disposition terms" (*Ibid.*, pp. 66-67).

<sup>18</sup> Carnap distinguishes pure dispositional predicates from theoretical terms by imposing the following requirement on terms of the first kind: "If *S* (stimulus) holds for *b*, but *R* (response) does not, then *b* cannot have the disposition." (*Ibid.*, p. 67). It should be noted in passing that this requirement would seem to reduce virtually every non-observational term to theoretical status; for the hypotheticals relating *S* and *R* terms will, in general, prove satisfactory only if they contain "*ceteris paribus*" clauses and this entails violation of the requirement for pure dispositional status.

<sup>19</sup> *Ibid.*, pp. 71-73. Carnap does not discuss beliefs and emotions directly but his point can be extrapolated to this case very easily.

alleged by Descartes and Ryle alike to be non-observational, are dispositional follows as a corollary. But then there is no need to lay special emphasis on the corollary. It is the general claim that wants defending, and that claim for reasons which are well known seems dubious. Hence the view that beliefs (and motives) are dispositions to action seems to be of doubtful significance when the aim is to clarify the conditions of applicability of non-observational terms.

However, that thesis may be serviceable for other purposes. We observed before that Ryle's concept of a disposition and the thesis that beliefs are dispositions to action indicates that he is concerned with another concept-formational issue pertaining to motive and belief predicates in addition to the one just considered. Since this issue has been formulated more clearly by Nelson Goodman than by anyone else, a consideration of Goodman's formulation seems in order.

According to Goodman, to attribute a disposition to an object is to assert that the object is "full of threats and promises." Dispositions, capacities, etc., "strike us as rather ethereal." We may therefore inquire as to whether "they can be brought down to earth, whether, that is, we can explain disposition terms without reference to occult powers."<sup>20</sup> Goodman continues:

The problem, then, is to explain how dispositional predicates can be assigned to things solely on the basis of actual occurrences, and yet in due accordance with scientific usage. In other words, what we want is a criterion in terms of actual occurrences—that is, in terms of manifest predicates—for the correct assignment of disposition predicates to things.<sup>21</sup>

Goodman's problem is a concept formational one. He wishes to specify conditions of applicability

of dispositional terms in a basic language which contains only manifest predicates where to apply a manifest predicate to a thing is "to say that something specific actually happens with respect to the thing in question."<sup>22</sup>

In the same context, Goodman implies that disposition predicates comprise all non-occurrent predicates. Indeed, almost every predicate commonly thought of as describing a lasting characteristic of a thing is as much a dispositional predicate as any other.<sup>23</sup> The basis for Goodman's assertion is that all non-occurrent predicates are full of threats and promises—i.e., speak of what can happen. Thus, for Goodman, the problem of disposition predicates is not that they are non-observable<sup>24</sup> but that they are full of threats and promises—a malady which infests all non-occurrent terms.<sup>25</sup>

Goodman's use of "dispositional" is more in accord with presystematic ideas than Ryle's in that he takes the conditions of applicability of disposition terms to be problematical. However, if, as before, we take Ryle's notion of a disposition predicate in a technical sense which construes assigning dispositional status as problem solving, not problem raising, there are grounds for saying that one problem Ryle felt could be solved is the problem of threats and promises. Belief predicates are non-occurrent and hence full of threats and promises unless some way of eliminating reference to what can happen can be found.<sup>26</sup>

According to Ryle, this can be done by construing belief predicates dispositionally—i.e., by taking belief attributions to be hypothetical statements or inference tickets. But, as Goodman recognizes,<sup>27</sup> this is an oversimplification. In the antecedent of the hypothetical, reference would have to be made not only to events but also to standing or boundary

<sup>20</sup> Nelson Goodman, *Fact, Fiction and Forecast* (Cambridge, Mass., 1955), p. 44.

<sup>21</sup> *Ibid.*, p. 45.

<sup>22</sup> *Loc. cit.*

<sup>23</sup> *Ibid.*, pp. 44-45.

<sup>24</sup> Goodman's concept formational problem is not to specify conditions of applicability of non-observational terms in some observation language. He is concerned to specify conditions of applicability of non-occurrent terms in an occurrent language. It should be noted, however, that Goodman, like Ryle, does not always distinguish clearly between the two problems.

<sup>25</sup> It is not clear whether Goodman wants to say that the threats and promises infect only non-occurrent terms. If he did, he would be mistaken. If "is magnetic" is full of threats and promises so is the occurrent, "becomes magnetic." Similarly with "is red" and "becomes red." Goodman admits that many *prima facie* occurrent terms speak of what can happen (although his reasons are different from that just cited) and concludes that the distinction between what is occurrent (i.e., manifest) and what is dispositional is context dependent (p. 59, footnote 7). But even if we concede this, some occurrent terms will be full of threats and promises; for, once again, if "is magnetic" is full of threats and promises in a given context, so is "becomes magnetic." This suggests that Goodman's program for solving the threats and promises mystery by specifying conditions of applicability of disposition predicates in terms of manifest (i.e., occurrent) predicates need not necessarily solve the problem.

<sup>26</sup> Evidence for this interpretation of Ryle can be found in his persistent conflation of the dispositional/non dispositional dichotomy with the non-occurrent-occurrent dichotomy. See G. Ryle, *op. cit.*, especially pp. 116-117. Note the title of chap. 5, "Dispositions and Occurrences."

<sup>27</sup> *Op. cit.*, pp. 17-24.

conditions which are specified in terms of non-occurrent predicates. Thus, soluble items dissolve when immersed in water provided that other conditions obtain throughout the process (e.g., the temperature of the water, its degree of saturation, etc.). Hence, either the unpacking must be continued until only occurrent terms appear—which neither Ryle nor Goodman nor anyone else has shown can be done—or some non-occurrent terms will have to be admitted to be non-problematical.<sup>28</sup> Since non-occurrent terms are, according to Goodman, full of threats and promises, some must not be problematical. Whatever the problem which dispositional analyses of belief are supposed to answer, it is not to be found solely in the fact that belief predicates are non-occurrent nor in the fact that they are full of threats and promises.<sup>29</sup>

Here lurks not only a general problem but also one for attempts like Ryle's which proffer dispositional explications of both belief and motives. And neither can the problem of standing conditions be avoided, for it is evident that what a man will do at a certain time is determined not only by his beliefs and not only by what happened to him (at or before that time), but also by the conditions under which he is prompted to act. Moreover, among these standing conditions, we shall include not only features in his environment but also his motives. In short, a specification of sufficient conditions<sup>30</sup> for an act *A* performed by Jones would include at the very least reference to his beliefs *B*, his motives *M*, the external conditions *E*, and the stimulus *S*. Consider additional acts *A'* and *A''* and the following sufficient conditions for them:

*M*, *B'*, *E*, and *S* for *A'*

*M'*, *B*, *E*, and *S* for *A''*

<sup>28</sup> This point applies to Goodman also. According to Goodman, we can make a beginning toward replacing disposition terms with manifest predicates in a manner illustrated by the replacement of "flexible" by "flexible-under-suitable-pressure-or-spectroscopic-inspection." This term applies to things that are either under suitable pressure and bend or are spectroscopically inspected. But what is suitable pressure? It must be pressure under suitable conditions including non-occurrent conditions. Goodman cannot escape this result by taking the hyphenated term as a single predicate; for even if, in some sense, it is occurrent, it is, like "becomes magnetic" full of threats and promises. (See footnote 25.) In all fairness to Goodman, it should be observed that the individuals in his discussion are not "long enduring objects but brief temporal segments of them." (p. 60, footnote 12.) It is not clear how Goodman means to distinguish manifest from dispositional predicates for such entities.

<sup>29</sup> This difficulty can be avoided by treating the pertinent distinctions as context dependent. We do not object to context dependent distinctions. But surely some indication must be given of those features of a context which determine how a given distinction is to be made in that context. From clues in the writings of both Goodman and Ryle, we suspect but cannot prove that their distinctions between dispositional and occurrent or dispositional and manifest predicates are determined by context in the same way (whatever that may be) in which distinctions between observational and non-observational terms are determined. We have already had our say about the relevance of such distinctions to dispositional analyses.

<sup>30</sup> By "sufficient conditions" we mean only that a statement of the following sort is true: Whenever anyone has beliefs *B*, Motives *M* in environment *E*, and is subject to stimulus *S*, he performs act *A*. It is doubtful whether we know any such statements to be true—certainly not without a *ceteris paribus* clause. But we shall conduct this discussion with the same simplifying assumptions that dispositional analysts allow themselves.

If we wish to take *B* dispositionally, attribution of *B* to Jones is tantamount to asserting (according to Ryle) that he would perform *A* under conditions *M*, *E*, and *S* and would perform *A''* under conditions *M'*, *E*, and *S*. *E* would contain reference to non-occurrences. Perhaps Ryle is not disturbed by publicly observable non-occurrences, but only by those which are non-observable. We have already had our say about the relevance of dispositional analyses of belief to their non-observational status and shall say no more here. But even if Ryle is taken to be concerned only with specifying conditions of applicability of non-observational non-occurrences, we still have to contend with the motives *M* and *M'*.

If *M* is construed as a disposition to *A* under conditions *B*, *E*, and *S* and to *A'* under *B'*, *E*, and *S*, we see that our analysis becomes circular. The circle can be avoided by giving sufficient conditions for action given a certain belief (motive) which do not contain reference to motives (beliefs). But it is highly doubtful as a matter of fact that sufficient conditions for human action can be given in terms of occurrences and publicly observable external conditions. If this factual claim is granted (and neither Ryle nor other contemporary dispositional analysts of psychological predicates seem to deny it), then neither unpacking belief predicates into hypotheticals, introducing them by reduction sentences, nor projecting manifest predicates a la Goodman will eliminate reference to non-occurrent terms and indeed to non-observational non-occurrent terms.

This point rests on the assumption that test conditions for psychological predicates cannot be specified in terms of predicates referring to external events or external events and standing conditions alone. Such a premiss itself automati-

cally implies that we cannot dispense with reference to "internal" non-occurrences. Ryle, Ayer, and other dispositional analysts of psychological predicates write as if treating such predicates dispositionally is a kind of logical magic which renders it possible to accept the assumption and avoid the implication. Our point is that the magic is impotent.

Our chief objective in this section has been to argue that analyses of beliefs as dispositions to action are of doubtful or at least minor relevance to concept formation issues. But we cannot conclude that the claim that beliefs are dispositions to act is without philosophical relevance. As we have already indicated many philosophers have insisted that the thesis is illuminating when we try to explicate "*A did X because he believed T*" or alternatively when we appeal to beliefs to explain human action. Once again let us begin by considering Ryle.

### III

As is well known Ryle's thesis, roughly speaking, is that to apply a disposition term to an object is tantamount to asserting what Hempel-Oppenheim call "an essentially but not purely generalized statement."<sup>31</sup> And hence to explain why a window pane broke by saying that it was brittle is not to indicate the conditions which occasioned the breakage. The application to belief is obvious enough, and with it the ostensible mistake of paramechanic theorists who consider belief predicates as characterizing causal factors for behavior rather than general principles governing behavior.<sup>32</sup>

Ryle's dichotomy is misleading, for one need not choose between construing beliefs as causal occurrent factors and disposing of beliefs as dispositions in Ryle's sense, i.e., principles. For among other things one could view belief-predicates as specifying standing conditions<sup>33</sup> in laws explaining human behavior, a view which

as we already indicated cannot be dismissed on the grounds that all non-occurent predicates which specify standing conditions can be eliminated.

Still Ryle might claim that some predicates which specify standing-conditions are legitimate and clear and others are not and require treatment as "open hypotheticals," and hence might grant that some non-occurent terms need not be eliminated even if they are in some appropriate sense dispositional. This is fair enough but not conclusive. For the problem now arises of distinguishing between non-occurent predicates which may be used to specify standing conditions and other predicates which need to be treated as "open hypotheticals." Moreover it would not be helpful simply to appeal to the meaning of "belief-predicates" for that appeal is in general inconclusive. In doing so we, among other things, beg the issue of the analytic and the synthetic, and what is worse forget that the decision to expand or to accept the expansion of a non-occurent term as a dispositional one in Ryle's sense is often based on theoretical and factual commitments.

A simple appeal to meaning being in general of little avail, two problems remain: the one of distinguishing between clear and unclear non-occurent terms that stand for standing conditions and the other of distinguishing between non-occurent terms that can be treated dispositionally and those that cannot. Here we might offer an aid to accomplish the latter task. If we can, with warrant, appeal in a causal explanation to a change in or of *T* (where "*T*" is a non-occurent term) as a or the cause of an event, then "*T*" can legitimately be viewed as involving a standing condition; if we cannot then "*T*" should be treated dispositionally in Ryle's and Carnap's manner. The motivation for this suggestion is clear. If disposition predicates cannot be used legitimately to specify standing conditions in explanations because they are law-like expressions, it would appear equally plausible to suppose that

<sup>31</sup> Ryle, *op. cit.*, pp. 123-124 and C. G. Hempel and P. Oppenheim, "The Logic of Scientific Explanation," reprinted in H. Feigl and M. Brodbeck, *Readings in Philosophy of Science*, pp. 343-344. Alan Donagan in "Explanation in History" (reprinted in P. Gardiner, *Theories of History*, pp. 428-443), contends that Rylean explanations differ "from anything recognized in Hempelian theory . . ." (p. 435). Ryle's law-like statements are exactly what Hempel and Oppenheim term essentially but not purely generalized statements, i.e., general statements which are not equivalent to singular statements but which contain individual constants. (The claim that they are not singular statements must, of course, be relativized to the set of predicates considered primitive within the context under scrutiny. Relative to such predicates, a singular statement is one which contains no variables). There is, indeed, a difference between Ryle and Hempel-Oppenheim. According to Ryle, law-like statements can be used in legitimate explanations even when they are not derived from purely generalized statements.

<sup>32</sup> Ryle, *op. cit.*, pp. 113-114.

<sup>33</sup> Ryle quite correctly takes the relation to obtain between events or occurrents.

change in disposition is, by assumption, a change in the regularities governing certain behavior and such changes cannot be causal factors. Hence, the same considerations that seem to account for the prohibition against using change in disposition as a causal factor also seem to make sense of a prohibition against using disposition-predicates in specifying standing conditions. And these considerations depend upon assuming that to attribute a disposition to something is to assert a law-like statement.

But if this criteria is acceptable, there is little comfort for Ryle and his explications. For consider "has temperature *K*". This predicate (when applied to liquid *A*) can be expanded into "if a mercury thermometer is placed in *A* under appropriate standing conditions, the mercury column will reach a certain level": and hence resembles "is butter," "is elastic," etc. To be sure, this expansion precludes using change in temperature in specifying a cause for a change in the height of the mercury column; but it does not prohibit appeal to change in temperature in specifying causal factors in explanations of other phenomena—e.g., in explaining why a given substance immersed in water for some time suddenly dissolves. In this respect, "has temperature *K*" differs from "is brittle."

We have reached an impasse. Given Ryle's and Carnap's proposal for treating terms like "has temperature *K*" and "is brittle" we have no clear way of distinguishing between accounts of the explanatory role of these terms. Here no issue is being begged. Ryle, Carnap, and others would agree that "has temperature *K*" is dispositional, and if they do not, they lack a way of characterizing the difference in the logical behavior of the latter term. And, of course, on Ryle's grounds we have no way of knowing whether we can appeal to a change of belief as a causal factor, and hence among other reasons the dubiety of his claim that beliefs are dispositions. And again if we insist that "has temperature *K*" describes a standing-condition we are left with the double problem of deciding on the one hand whether it is or is not a legitimate, non-puzzling term, and on the other hand, accounting for its ostensible treatment as a dispositional term in Ryle's sense.

Of course in offering this argument we are not forgetting one previous objection, that a given

non-occurrent term is not always obviously expandable in virtue of its meaning into a Rylean hypothetical, and that the expansions offered in the case of determinable disposition-predicates are often based on empirical or theoretical considerations rather than on meaning. Moreover, even if we concede the distinction between the analytic and the synthetic, we must also admit that there is no unique way of dividing true statements into these two categories. Statements which assert that having a given disposition implies the truth of certain Rylean hypotheticals may in some contexts be taken to be analytic and in others to be synthetic without in any way altering the explanatory role of the disposition predicate or predicates characterizing change in disposition. But if it is not necessary that disposition predicates imply hypotheticals in virtue of meaning but perhaps only for empirical or theoretical reasons, we can say the same for any extralogical predicate, either occurrent or non-occurrent.

Nevertheless, we feel that assigning belief-predicates dispositional status is a significant thesis in the context of explanation. But to avoid the previous snares we cannot simply follow Ryle and Carnap in their treatment of disposition and must seek a new road. Interestingly enough, a clue as to the direction we ought to follow is offered by Ryle, himself, who acknowledges that in one sense the notion of dispositionality is mystery-raising not mystery-solving.<sup>34</sup> "The old error of treating the term 'Force' as denoting an occult force-exerting agency" is taking "Force" to be dispositional in the mystery-making sense. Dormitive powers, cognitive faculties, etc., are also mystery-making. To take belief-predicates to be dispositional is, within the context of explanation, to claim that explanations of behavior in terms of beliefs are in some sense mysterious. The next step is to indicate what the mystery might be and what steps must be taken to solve it.

In the brief, but illuminating, discussion of disposition terms in *Word and Object*, Quine notes that the use of disposition terms differs only in degree from the use of "*caeteris paribus*" clauses in generalizations.<sup>35</sup> Why are *caeteris paribus* clauses ever used in explanations? The most obvious suggestion is that the insertion of such clauses is an attempt to convert generalizations which are at best essentially but not purely generalized into

<sup>34</sup> *Op. cit.*, pp. 119–120. We have already noted that Goodman also construes dispositionality to be problematic or mystery-raising. We feel, however, that he (and Ryle when he talks this way) fail to identify the mystery.

<sup>35</sup> W. V. Quine, *Word and Object* (New York, Technology Press and Wiley, 1960), p. 225.

pure generalizations. *Ceteris paribus* clauses are devices for fitting explanations into a Hempel-Oppenheim procrustean bed. They are place holders for unspecified standing conditions in purely general laws.

We do not mean to imply that explanations, using essentially but not purely generalized statements which are not derived from fundamental laws (i.e., Rylean explanations), are not genuine explanations. However, they are extendable in the sense that it would be legitimate to ask for additional explanation. This contrasts with the ideal case of non-extendable explanations where (if we had them) the demand for additional explanation would be illegitimate—i.e., would be like asking why the world is like it is. The position taken here is that conditions like Hempel-Oppenheim conditions are necessary (but not sufficient) for non-extendable explanations. The introduction of *ceteris paribus* clauses is an acknowledgment that these necessary conditions are not fully satisfied. To repeat, they are place holders of unspecified standing conditions which when found will reduce the degree of extendability of explanations in which laws containing such clauses appear.

Disposition predicates, like *ceteris paribus* clauses, function as place-holders for predicates specifying conditions in generalized statements. But they are not simply place-holders and differ from *ceteris paribus* clauses in a number of respects: *Ceteris paribus* clauses entail no commitment as to the kinds of predicates to be employed in replacing them; disposition predicates do. Hence, a given disposition predicate may be a surrogate for standing conditions in many different laws; and unlike *ceteris paribus* clauses, this involves a commitment (which may prove erroneous but which may have at the moment some inductive backing) to replacing the disposition predicate by the same specification of standing conditions in each of the laws in which the disposition predicate appears. Thus, if magnetism is a disposition, it is a disposition to attract and repel iron filings

and to induce electricity in electric coils. The reference to magnetism rather than the use of a *ceteris paribus* clause in explaining these two phenomena implies that a certain standing condition is the same in explanations of both phenomena.<sup>36</sup>

To further sharpen our formulation, we note that whereas *ceteris paribus* clauses are promissory notes for meeting Hempel-Oppenheim requirements, disposition terms sometimes entail a commitment to look for standing conditions which meet other necessary conditions for non-extendable explanations. Thus, if we are told that objects break when tapped lightly *ceteris paribus*, we are told very little about how the blatant deficiencies of this generalization as a law can be eliminated, save that Hempel-Oppenheim requirements are to be met. We do not know what kinds of predicates are to be used in specifying adequate standing conditions. For all we know they may refer to observable conditions in the environment, observable features of the windows or the micro-structures of the windows. On the other hand, if we are told that fragile objects break when tapped lightly, we assume that if we are to improve or replace the generalization we should investigate the micro-structure of fragile objects.<sup>37</sup>

Hence *ceteris paribus* clauses are *ceteris paribus* clauses in every context in which they occur, whereas a given term is taken to be dispositional when it is considered theoretically unacceptable for use in specifying standing conditions. Thus, whether notions of utility and personal probability, when used in explanations of human choices, are taken to be dispositional or not depends not only on the empirical adequacy of the decision theory employed but also upon whether a demand is made that explanations of human choices be stateable in terms of a neurophysiological (or psychoanalytical or some other) theory or whether the use of terms from decision theory are considered adequate in non-extendable explanations.<sup>38</sup>

Allowing beliefs to be dispositions we therefore

<sup>36</sup> The commitment to look for the same standing conditions in all contexts in which a given disposition term is used may be abandoned without abandoning the use of the disposition term in all these contexts. If we have good evidence to the contrary, we need not suppose that "fragile" is to be replaced by the same specification of standing conditions in every context in which "fragile" is used as a place holder. However, the use of the same disposition term in explanations of different phenomena does involve a *prima facie* commitment to look for the same or similar standing conditions.

<sup>37</sup> We do not mean to require that disposition terms be replaced by specifications of micro-structure. The type of characterization of standing condition which is suitable as a replacement of a disposition term depends on context in the sense indicated in the next paragraph and footnote 38.

<sup>38</sup> Thus, we agree with both Goodman and Quine that the distinction between dispositional and non-dispositional terms is context dependent but disagree with both of them as to how the context determines the distinction. According to Goodman the

treat "belief" as a certain kind of place-holder term. But still enough has not been said to allow us to answer some of the questions we have raised, for example, whether change in belief can legitimately be considered a causal factor. To deal with this and other problems we must expand and distinguish between types of dispositional terms, partly but not completely along lines indicated by others.

C. D. Broad and C. L. Stevenson, among others, have observed that dispositions are frequently associated with what Stevenson calls "bases."<sup>39</sup> According to Broad, the bases of all physical dispositions are microstructures of the objects of which the dispositions are predicated.<sup>40</sup> Prior to the rise of modern physics, Broad contends, physical dispositions were not associated with microstructural bases. Medieval physics was therefore a "faculty physics."<sup>41</sup> Psychological dispositions continue to lack bases analogous to those which can be supplied for physical dispositions. Broad concludes "that we cannot get away from the much decried faculty psychology."<sup>42</sup>

According to the "place-holder" conception of disposition predicates we are proposing, not every disposition need have a basis. The basis (if any) of a disposition is the set of conditions which are specified by that description which we are entitled on theoretical and empirical grounds to substitute for the disposition predicate. Whether the basis must be given in terms of a microstructural description and if so what kind is relative to the set of predicates to which we are committed by our theories in the sense discussed before. An example of a basis which is not microstructural is one in which the force exerted on a body is specified in terms of relations between that body and features of its environment by means of force functions. Since we do not always have a legitimate characterization of conditions to take the place of

disposition terms, we cannot always associate a disposition with a basis.

Hence it seems best to distinguish between three kinds of dispositional-predicates; those for which legitimate bases have been found, those lacking them though allegedly not, and those both lacking bases and claim for such bases. And whereas Broad insists that only dispositional-predicates of the first type avoid reference to faculties and considers the last two jointly, it seems best to distinguish between them and consider them separately. For obvious reasons the third kind can best be viewed as problem-raising, the second mystery-making, and the first mystery- and problem-solving. Now having distinguished between them, we can offer dividends and try to answer some questions previously left begging for treatment.

It will be recalled that Ryle and others had no clear way of distinguishing between those changes in dispositions which can and those which cannot be considered causally. Now we might suggest that the first type of dispositional term can and the latter cannot be so considered. The first is associated with a base, changes of which can be viewed as an event. But the other two lack such bases, and are, as it were, mere place-holders, and hence, here, changes in dispositions are not causes and by the same token are not effects.

Broad has noted that dispositions are frequently of different orders. Magnetizable objects have a disposition to become magnetic when certain operations are performed on them; magnetizability being, therefore, a higher order disposition than magnetism. It is clear that in using "magnetizable" we presuppose that "becomes magnetic" can be used legitimately to describe an effect. But, according to the view here defended, this implies that "magnetic" has a basis, that it is a disposition of the first kind, namely, those that have legitimate bases. Other-

distinction is relative to the basic language of occurrences; according to Quine, a predicate is dispositional relative to what it is dispositional to; according to our proposal, it is relative to the set of predicates which can be used to specify a legitimate replacement of the disposition predicate *qua* place-holder.

<sup>39</sup> C. D. Broad, *Mind and Its Place in Nature* (New York, Harcourt Brace, 1925), pp. 430-440; C. S. Stevenson, *Ethics and Language* (New Haven, Yale University Press, 1940), pp. 46-53.

<sup>40</sup> *Op. cit.*, pp. 433-435.

<sup>41</sup> *Ibid.*, p. 440. Broad did not maintain that physics could eliminate all dispositions ("causal characteristics") by micro-reduction. To characterize the microentities as obeying certain laws is, so Broad contended, to ascribe to them certain causal properties. If these laws are purely and essentially general, we do not feel that any purpose is served in talking of the entities obeying them having a disposition. According to our "place-holder view" the point of introducing disposition predicates is to reduce the extendability of explanations. When the available laws provide us with non-extendable explanations, there is no reason for talking of a disposition to obey these laws. Of course, we are not saying that microentities must lack dispositions. We may not have purely and essentially general laws expressed with theoretically impeccable predicates. Moreover, what counts as purely and essentially general depends on the range of our individual variables.

<sup>42</sup> *Loc. cit.*

wise "magnetic" is a mystery-making disposition.

It is therefore not surprising that physicists frequently avoid and psychologists do not avoid the introduction of occult qualities or faculties. For in many areas of physics where bases for disposition terms can often be supplied on demand or the state of our knowledge warrants our expectation that further routine research would provide us with the appropriate bases, the dispositions are of the first, the clearest type. In psychology, there is no consensus regarding the kinds of predicates to be used in specifying bases for psychological dispositions, and there are very few grounds at present for expecting that bases satisfactory from the standpoint of current theoretical commitments can be found without an important scientific breakthrough. Nevertheless, we talk about acquiring and shedding psychological dispositions and frequently make reference to higher order dispositions like irascibility, and therefore often consider changes in dispositions effects. But in doing so, we are introducing predicates which are mystery-making, and leave ourselves open to Broad's contention that psychology remains faculty psychology.

Whether this severe judgment is in order when we appeal to beliefs to explain human action cannot be answered simply. To be sure, that common sense takes changes in belief to be effects can be seen in everyday accounts of persuasion. That changes in belief are alleged to be causal factors is exemplified in the common sensical reference to changes in Kennedy's beliefs (about the kinds of arms being shipped to Cuba) in explaining his decision to impose a blockade on Cuba. But if our previous observations are in order, dispositional analysts of belief must either refuse to permit common sense practice, supply bases for belief-dispositions (or at least indicate the direction which will lead to their discovery), admit that belief dispositions are for the present mystery-making in a sense akin to that in which dormitive powers are, or abandon the claim that beliefs are dispositions to action. Our complaint against Ryle, Ayer, *et al.*, is that they are not prepared to do any of these things.

Of course, Ryle and others are quite correct when they insist that on occasion we can in a way

explain why *A* did *x*, by first insisting that *A* has a disposition *T*, and then expand the disposition in a familiar manner. But this is a half-way house. To say *A* yells because he is irritable should lead us to say not merely that *A* yells under certain conditions but that all irritable men yell under them. And it is the use of "irritable" in the latter type of statement that Ryle and others do not illuminate, and the place-holder view, we hope, does.

There are, we trust, additional points in its favor; alternative accounts of this distinction between dispositional and non-dispositional analyses of belief are often untenable, misleading, or of little philosophic or scientific interest, while the place-holder account clarifies the sense in which some dispositions are mystery-making and indicates how the mystery can be solved. It does so in a manner that provides a place for the fact that dispositions are often alleged to have bases and that having a basis renders a disposition less mysterious. (Thus, the place-holder view is not entirely arbitrary. Indeed, it seems to cope with some presystematic notions of dispositionality better than its competitors).

None the less, it may be argued that the place-holder account is counterintuitive because it implies that changes in dispositions cannot be effects, and hence, that explanations of the acquisition of first order dispositions by reference to second order dispositions is illegitimate. But this charge is unfounded; the place-holder account does not say that change in disposition cannot be an effect save when the disposition lacks an adequate basis. If this claim were counterintuitive (it is not obvious that it is), we would be compelled to convict our presystematic ideas of inconsistency, for we usually assume that what can be an effect can be a cause. But since changes in elasticity, brittleness, solubility, etc., are not considered causal factors in explanations of other changes,<sup>43</sup> our refusal to countenance them as effects seems appropriate. We readily concede that this might in some cases call for altering presystematic usage, but the change here might very well be an improvement and avoid the violation of the symmetry requirement for causes and effects which may lurk in presystematic usage.

But despite these virtues we present the place-

<sup>43</sup> We may in a certain sense explain why a metal ring did not regain its shape after it was squeezed by reference to its loss of elasticity. But the context in which such explanations occur is one where it was initially assumed that the ring was elastic and that it would regain its shape. Our request for an explanation is, in this case, a request for the removal of an inconsistency. The reference to loss of elasticity is not, therefore, to be construed as a specification of a causal factor but as an indication of the way in which the inconsistency is to be removed.

holder view as a proposal for making the distinction between dispositions and non-dispositions. Moreover, we want to re-emphasize the point that the applications of the proposal for the elucidation of specific terms and phrases is not mechanical and depends essentially upon scientific as well as philosophical decisions. The last must be applied to "belief." The thesis that beliefs are dispositions to act is a claim about the role of belief predicates in explanations of human behavior; it alleges that belief predicates are theoretically defective for that purpose. The tenability of this thesis depends on the theories in which belief predicates appear, the empirical adequacy of such theories, and their explanatory or, more generally, their systematic power. It also depends upon commitments as to the kind of theory (if any) to which psychological explanations of human behavior are to be reduced.

The extent to which criteria can be offered for empirical adequacy, systematic power, and basic (with respect to theoretical reduction) scientific disciplines remains undetermined. For this reason, we are not prepared to insist upon taking belief-predicates to be mystery-making dispositional terms or (alternatively) upon construing them as specifying theoretical standing conditions in explanations of human behavior. The important point is that adoption of the former alternative is a criticism—not a defense—of ordinary ways of using belief-predicates. Such criticism may very well be in order. What is out of order is the assumption that mysteries-surrounding, paramechanical legends are thereby solved. In sum, to say that beliefs are dispositions to act is not to solve a philosophic problem but to raise a scientific one.

*The City College and Columbia University, New York*

## V. REFERENTIAL INDENTIFIERS

K. W. RANKIN

"But surely the word 'I' in the mouth of a man refers to the man who says it; it points to himself: and very often a man who says it actually points to himself with his finger." But it would be quite superfluous to point to himself. He might just as well only have raised his hand. It would be wrong to say that when someone points to the sun with his hand he is pointing both to the sun and himself because it is *he* who points; on the other hand, he may by pointing attract attention both to the sun and himself. . . . The word "I" does not mean the same as "L.W." even if I am L.W. nor does it mean the same as the expression "the person who is now speaking." (Wittgenstein, *The Blue Book* (Oxford, 1958), p. 67.)

RUSSELL refers to personal pronouns (e.g., "I," "he," "you") and demonstratives (e.g., "here," "now," "this") as egocentric particulars<sup>1</sup> to signify, it would seem, that they are all inter-definable. For convenience he defines them in terms of "this" but also shows how the composite "I-now" could do the job as well (or as badly). Reichenbach in his *Elements of Symbolic Logic* (N.Y., 1948) offers a more detailed version of this view, and this version has had the greater impact upon subsequent philosophy mainly by virtue of his characterization of these expressions as token-reflexive and his further detailed application of this characterization to tensed statements.<sup>2</sup> I propose to show in the next five sections that while expressions of this sort do, in a somewhat attenuated sense, function in a token-reflexive manner any contention that they actually refer to their own tokens is in the main mistaken. In the last four sections I investigate wider and more positive implications of this conclusion.

### I. DEFINITIONS OF

#### "I," "NOW," AND "THIS TABLE"

In what has turned out to be one of the more exoteric parts of his analysis Reichenbach states

It is easily seen that all these words can be defined in terms of the phrase "this token." The word "I," for instance, means the same as "the person who utters this token"; "now" means the same as "the time at which this token is uttered"; "this table" means the same as "the table pointed to by a gesture accompanying this token." We therefore need inquire only into the meaning of the phrase "this token."<sup>3</sup>

This justification for selecting the phrase "this token" for particular scrutiny is less acceptable than he thinks. His definitions would seem to be offered as of the explicit sort where the definition can be substituted for what is defined. But where two equivalent expressions conforming to this condition are referring expressions, as those above are, they must refer to the same thing. Otherwise they cannot be equivalent. For example, "the neighbors" and "the people who live in the immediate vicinity" are interchangeable only to the extent to which they refer to the same thing or things.

Reichenbach's definitions do not fulfil this condition. In that of "I," for instance, to what does "this token" refer in the phrase "the person who utters this token"? It cannot refer to a token of "I" because when I use the definiens-expression no appropriate token of "I" has been uttered. If "this token" had here been intended to refer to a token of "I" the definiens-expression couldn't be used at all to make a true or false statement since "this token" would no longer have an application.

It is just this interpretation, however, that seems to me to lend Reichenbach's definitions an initial plausibility. Although "this token" cannot refer to a token of "I" in his definition, one could correctly say that "I" is so used that *listeners* or *readers* can understand by it "the person who utters this token." Here we have a definition in use of what one person means by the word in terms of what *other people* can say when reporting him or he himself on a different occasion.

<sup>1</sup> *An Inquiry into Meaning and Truth* (London, Allen & Unwin, 1948), chap. 7, and *Human Knowledge*, pt., II, chap. 4. In *The Philosophy of Logical Atomism* he calls them "emphatic particulars" (see *Logic and Knowledge*, ed. by Robert C. Marsh [London, Allen & Unwin, 1955], p. 222).

<sup>2</sup> For a recent testimonial see J. J. C. Smart, "Tensed Statements': A Comment," *The Philosophical Quarterly*, vol. 12, (1962), p. 264.

<sup>3</sup> *Op. cit.*, p. 284.

But it is the alternative interpretation that Reichenbach favors. The phrase "this token" might be taken to refer to the token of the translation (e.g., "The person who utters this token"), instead of to the token of the expression translated (e.g., "I"). It would apparently then enjoy a built-in guarantee that it was a referring expression which applies to something on every occasion of its utterance. But this guarantee has only been issued at the cost of making the translation refer to something to which what it translates cannot possibly refer. If *A* says "I am overweight" he is not uttering the translation-token "The person who utters this token is overweight." Hence if the latter referred to itself it cannot possibly be implied by *A*'s statement "I am overweight," and surely that is enough to prevent it from being a genuine translation.

All these criticisms apply alike to all of the three definitions offered by Reichenbach in the above passage. In addition, however, his definition of "this" raises difficulties peculiar to itself. The difference between his and Russell's exposition seems marked by his having noticed that "this" is an adjective. It qualifies an indefinite variety of nouns though sometimes these are understood or left indefinite.

However, if it is possible to give a definition in use of "this" as in "this table" in terms of "this token," it should also be possible to give a definition in use of "this" in "this token" in terms of "this token" along similar lines. The phrase "this token" would have to mean the same as "the token indicated by a gesture accompanying this token," and so on in vicious regress.

To escape the difficulties peculiar to his analysis of "this" Reichenbach would have to stipulate that its function in the phrases "this token" and "this table" are not the same. He might say that the contrast between the phrases "this table" and "that table" is quite different from the contrast between "this token" and "that token." In the former couple both the "this" and the "that" refer to some object apart from themselves, whereas in the latter pair he might say that only the "that" points beyond itself, for the "this" refers to itself. The function of the word "token" could then be said to be simply that of signalizing this disparity.

We might elaborate this way out further in terms of Wittgenstein's analogy (see above). Reichenbach's definition of "this table" is tantamount to the supposition that the man who points

at the sun points to himself pointing to the sun as well, and this would seem to lead regressively to the supposition that the man in pointing to himself pointing to the sun is also pointing to himself pointing to himself pointing to the sun. At this point, however, one could resolve to check the regress by decreeing that pointing to oneself is a special fundamental sort of pointing which provides the bases for the other pointing but itself needs no further basis.

To follow this line is to engage, however, in replacement as distinct from exhibition analysis for two reasons. First, in so far as ordinary discourse offers any guidance to the logic of "this token" it is by analogy to expressions like "this table." Hence it offers no justification for singling out "this token" from the other "this"-phrases as the point at which the regress must be checked. Ordinary usage offers us the choice here between continuing the regress or abandoning the attempt to analyze the other "this"-phrases in terms of "this token." Second, as we have seen, in making "this token" refer *inter alia* to itself, Reichenbach has ensured that the translations in which it occurs must have a different reference from the ordinary expressions which the translations are supposed to translate.

Once these consequences are pointed out, the only motive for continuing the analysis must lie in a desire to get behind the presuppositions of ordinary language. Communication, as John MacMurray has emphasized,<sup>4</sup> rests upon three presuppositions, viz., the existence of the person making the communication, the existence of somebody with whom he communicates, and the existence of a subject-matter about which they communicate. Reichenbach's analysis would appear, intentionally or otherwise, to attempt to reduce the first two to the third, and at the same time through its reflexivity to provide a built-in guarantee that the third presupposition is fulfilled once this reduction has been carried out. We must now see whether such a project can hope to succeed by examining the special rôle which he has devised for "this token."

## II. TOKEN QUOTES

He interprets the phrase "this token" as an operation with a function similar to that of ordinary quotation marks, i.e., as a token-quotes operation.

<sup>4</sup> See "Some Reflections on the Analysis of Language," *The Philosophical Quarterly*, vol. 1 (1951), p. 324.

Whereas the ordinary-quotes operation leads from a word to the name of that word, the token-quotes operation leads from a token to a token denoting that token.<sup>5</sup>

He then invents an arrow sign for the token-quotes operation and proceeds to describe any application of this operation as reflexive, apparently because the token resulting from the operation, i.e., the arrow-quoting token, physically includes the token which it is supposed to denote, i.e., the arrow-quoted token.

A further feature which the arrow-quoting token acquires from the arrow-quoted token is that it cannot be repeated without loss of identity. The arrow-quoting token can, however, be named—e.g., by numerals—and these names can be repeated *ad lib.* for referring purposes. Let (1), for example, be the statement saying that an inscription of the word "cow" (duly put into arrow quotes) is legible. Then (1), though markedly unlike Reichenbach's own application of the arrow-quotes operation, would appear to fit his specifications for that operation exactly. More specifically, the subject of (1) if it refers to anything at all refers to *part* of itself, and it possesses no equisignificant replicas.

Now I shall argue that this account of the token-quotes operation doesn't pay enough attention to a distinction between two ways of referring to tokens. The first rests on the assumption that the tokens referred to have been provided prior to the referring except in certain limiting cases which we shall investigate. The second, which approximates more closely to Reichenbach's token quotes, provides or attempts to provide the tokens in the act of referring to them.

### III. TOKEN-REFERRING: FIRST METHOD

To reveal my hand at once, the first method commonly makes use of phrases like "this token," or more colloquially "this word," "this sentence," etc., where the context makes clear that one is referring to tokens rather than types. One could also substitute the word "that" for "this" for the

purpose of referring to tokens which might, but need not, be more distant. I could, for instance, point to a token in the previous sentence and say "This (token) is illegible" or "That (token) is illegible." Here of course there is no question of the tokens referred to being quoted—at least not through the act of referring. They cannot be quoted in the act of referring because, except in certain limiting cases, they are tokens which have already been formed, and a quotation of these tokens could only produce different tokens.

There are, however, two ways in which phrases like "this token" can be used for referring by the first method—and this brings us to the limiting cases. They have a reflexive use, to refer *inter alia* to one of their own tokens as well as the non-reflexive use illustrated above. I might say, for instance, "The first token in the previous sentence is illegible." Here the phrase "the first token" may appear to refer non-reflexively, but it can refer unambiguously only because it is anchored to the phrase "the previous sentence," and this of course means the same as "the sentence before this sentence." Here then I was obviously making a reflexive use in an indirect way of a "this"-phrase. Such perfectly normal locating devices as "See above" and "See below" do this as well. Compare also the sentence "This is middle C" intoned in middle C as an effective demonstration of middle C.<sup>6</sup>

Reflexive uses of this sort, however, are just extensions of the non-reflexive uses of the same referring expressions. They are not endowed with the *essential* type of reflexivity which Reichenbach's token-quotes operation confers upon an expression. No token-quoting token can be "denotatively" equisignificant to other token-quoting tokens of the same form or sound. In fact because of this Reichenbach (p. 286) calls the phrase "this token," which he wrongly believes to perform the token-quotes rôle, a pseudo-phrase on the rather outmoded assumption that the meaning of a type is fixed by its denotation.<sup>7</sup> But quite to the contrary, of course, the phrase "this token," etc., is frequently used on

<sup>5</sup> *Ibid.*

<sup>6</sup> C. W. K. Mundle's illustration of token-reflexivity: see his "Broad's Views About Time," *The Philosophy of C. D. Broad*, ed. Schilpp (N.Y., Tudor, 1959), p. 370.

<sup>7</sup> The same assumption seems to lead Goodman (in *The Structure of Appearance* [Cambridge, Mass., Harvard University Press, 1951], p. 288) to say that the *tokens*, not the *types* of what he calls "indicators," function as words or sentences, while adding that the term "token" in this context is no longer appropriate. This in turn leads him (p. 290) to view the fact that very often the replicas of such words ("tokens") "name" different things as an extreme case of the ambiguity possessed by such names as "Cambridge" or "Paris."

different occasions to denote exactly the same thing.

For example, I could now, if I wished, refer to the previous sentence as "this token" for an indefinite number of times. Or I could make a statement in which I refer reflexively to that statement by means of "this token," and then go on to make statements about that statement still using "this token," this time non-reflexively but in exactly the same sense and with exactly the same reference. Hence any reflexivity in the use of "this token" is *accidental* to its sense.

It is in fact only because this reflexivity is accidental, or a limiting case of a non-reflexive use, that its use can make sense. The mere fact that self-referring expressions of this sort have an honorable rôle in our language should not blind us to the fact that in some ways they are viciously circular. All it shows is that the effects of the vicious circularity have in some way been circumvented or neutralized.

Certain Gaelic place-names, for instance, have circular forms when translated, such as "The Loch of the Mountain of the Loch." Taken at its face value as a recipe for locating a certain geographical feature the apparent circularity here would render the recipe quite useless. But of course the Gaelic name probably never functioned as a recipe for locating and always as a name. In any case it works well enough for the purpose of distinguishing one loch from another in the way that names distinguish where we are already acquainted with its location or can become acquainted by independent means.

Similarly an intoned token of the sentence "This token is middle C" if used to make a statement might be viciously circular in a way, but not a similar way, that can be ignored. Before the sentence-type can be used as a statement the phrase "this token" must have an application, for it presupposes the existence of a token; but before "this token" can have an application, considered as a *token in use* as distinct from a mentioned token, the sentence-type must be used for the purpose of making a statement.

This viciousness, however, is neutralized in two ways. (a) The spatio-temporal criteria, which permit a token of the same sentence to be used externally (i.e., a short time before or after the intoned token or in the same vicinity) to refer to

the intoned token, also permit spatio-temporal identity with the intoned token as a limiting case. (b) The word "this" when used by one person is readily translatable into "that" as used by another or the same person at a time or place more remote from the original subject-matter. When *A*, pointing to an exhibit in a display-case, says "This is Baron von Richthofen's flying-boot," *B* can and in some situations must translate what *A* says as "That is Baron von Richthofen's flying-boot." Similarly when *A* says reflexively "This sentence is reflexive,"<sup>8</sup> *B* can translate into "That sentence is reflexive," which is non-reflexive. In short, token-reflexive sentences are non-viciously circular in either of these two ways through acquiring what I shall call an "alienated use."

As a final point we must also note that in its alienated use "this token" can and, so far as I can see, must refer to the whole of itself. It is wholly reflexive as well as accidentally reflexive.

#### IV. TOKEN-REFERRING: SECOND METHOD

If we are to regard the token-quotes operation as a referring device at all, its distinctness from the first token-referring method even in its alienated use is now quite apparent. Unlike the latter it is essentially but only partially reflexive, if it is reflexive at all. I shall now argue that it isn't reflexive at all.

Expressions in quotation marks are pictures, or, as the later Wittgenstein remarked, patterns<sup>9</sup> or samples.<sup>10</sup> Ordinary quotes, for instance, operate mainly in two ways. (a) They make use of a given token in order to mention a replica or replicas of that token as used on a specific occasion or set of occasions to state, curse, etc. (e.g., "He said 'Go to hell'"). (b) As used by grammarians, metalogicians, etc., they mention a type as used in any of the actual or possible replicas of the token of that type within the quotes (e.g., "'Go to hell' is a curse").

Token quotes would appear to be a limiting case of quotes, and a limiting case just because they are a limiting case of picturing. We could account for the possibility of the operation in the following way.

In origin at least, pictures are representational, but even where they remain representational they come to be viewed for their intrinsic aesthetic

<sup>8</sup> I owe the example to Eric Toms.

<sup>9</sup> *Brown Book* (Oxford, Blackwell, 1958), p. 84.

<sup>10</sup> *Philosophical Investigations* (Oxford, Blackwell, 1953), § 16, p. 7.

value, and because of this completely non-representational forms of art develop in the course of time. Suppose we take as one of the characteristics which establish a picture as a picture its discontinuity with its physical context. We might then establish the convention that when we view a picture as representational, even when what it depicts does not exist, we put it into an ornamental frame; and if we want to view it non-representationally as an object in its own right we use a plain frame instead.

In these circumstances ornamental frames would be the analogue of ordinary quotes. In ornamental frames physically similar pictures would be taken as all depicting the same scene, person, etc., and similarly in ordinary quotes physically similar tokens are taken to be tokens of the same word. But in plain frames physically similar pictures would be viewed in their own right without reference to each other or to anything else, and similarly in the case of physically similar token-quoted tokens.

But despite these differences, or perhaps to a greater degree because of them, the token-quoting type of picturing is subject to one of the limitations of the more paradigmatic representational type of picturing. The token-quoting type of picture cannot depict itself, (1) because it is not a depicting picture at all, and (2) because in any case depicting pictures cannot depict themselves.

There are two quite familiar models for transcending this limitation, (a) Bradley's example of the label on the beer bottle which tries to depict the beer bottle, label and all, and (b) the line which forms itself into a picture of the artist or team of artists drawing it, as in some of Steinberg's "All in Line" drawings. In the first case the picture tries to depict the whole of itself as part of itself and gives up at some point well short of the necessary infinite regress. In the second case the picture tries to depict the whole of itself as the whole of itself, but faces the dilemma of either remaining incomplete or else becoming complete at the expense of becoming only a picture, say, of a man scratching his nose with the nib of his pen. These models indicate logical limitations to picturing. A condition of the picturing relation is the discontinuity between itself and its environment and this relation *ex hypothesi* cannot be incorporated in the picture.

I conclude that the token-quotes operation has no bearing upon token-reflexivity whatever.

<sup>11</sup> *Ibid.*, p. 286.

<sup>12</sup> *Ibid.*, p. 286.

## V. TOKEN QUOTES AND "THIS TOKEN"

Reichenbach is not entirely insensitive to the distinction between the functions of "this token" and token quotes, but he tries, with disastrous results, to dismiss it as one of complexity. He observes in a revealing footnote

There is a slight difference in the usage of the phrase "this token" and the usage of arrow quotes. The token including the arrow quotes refers to the same token with the exclusion of the arrow quotes, whereas the phrase "this token" refers to the latter token including the token of the word "this." The tokens of the word "this" are therefore reflexive whereas the arrow quotes are not. This difference is irrelevant.<sup>11</sup>

Whether or not this admission goes far enough the need for it is fairly obvious. If a token of "this token" did not refer to itself as well as to the token which accompanies it we could give no sense to

the table indicated by a gesture accompanying this token (2)

unless "this token" is here used non-reflexively to refer to some token in the vicinity of (2) but external to it. We should be compelled to erase the token of "this token" from (2) and to place the residue in arrow quotes, producing thereby a token similar to the specific inscription:

the table indicated by a gesture accompanying (3)

But (3) is not a complete referring expression, whereas (2) would seem to be.

What this incompleteness indicates, however, is not that "this token" is a more comprehensive sort of token-quotes operator which operates *inter alia* on itself—for even a representational picture cannot picture its frame-device—but rather it *refers inter alia* to itself by the method outlined in section 3. Reichenbach's own illustration of the way in which "this token" works as a quotes-operator would appear to make this more evident, for he says:

the piece of paper covered by this token (7)

could be replaced by a sentence [*sic*] consisting of the token (7) in token quotes, and the words "the piece of paper covered by" written before it. Or using the name "(7)," we could write on any sheet of paper the sentence [*sic*] "the piece of paper covered by the token (7)." In this form the token-reflexivity has been eliminated.<sup>12</sup>

Now it would seem fair<sup>13</sup> to suppose that Reichenbach means that only the last of these two translations eliminates token-reflexivity. Otherwise he would be saying that the arrow-quotes operation which by his definition constructs a reflexive token could *per impossibile* be used for eliminating it. Even so the first translation doesn't seem to tally with his original account of arrow-quotes.

To recapitulate, he has defined the arrow-quotes operation as leading to token-reflexivity by begetting a token which denotes a token which is part of itself. In the first translation of his (7), however, the operation is represented not merely as begetting a token which denotes what it includes, but as a method for constructing a still more inclusive phrase-token which

- A. (i) refers to a piece of paper
- (ii) refers to a token-quoted token
- (iii) identifies the first by referring it to the second.

Now granted that any quotes, token or ordinary, must be part of a wider expression, yet there is nothing in Reichenbach's original definition of token quotes to dictate what the wider expression would be. His (7) could, for instance, have equally well formed part of a phrase, introduced by the "piece of blackboard covered by" or "The person who utters," etc.

In short, the token-quotes operation was devised solely for the purpose of function (ii) above, whereas Reichenbach's illustration invests it with (i) and (iii) as well. But can this accretion to its functions for the purpose of interpreting the use of "this token" fairly be dismissed as "a slight difference"? Reichenbach appears to have thought that he was only making the requirements of function (ii) slightly more stringent by adding to them that the token-quoting token and the token-quoted token should coincide. Instead he seems to have amalgamated a token-referring expression with a token-quoting expression.

As a token-referring expression his (7) would

- B. (i) refer to a piece of paper
- (ii) refer to its *own* token
- (iii) identify the first by referring it to the second.

As a token-quoting expression on the other hand it would cease to do anything like (i) and (iii) and

instead of (ii) it would quote its own token. Of the two functions, therefore, the token-referring function would come closest to producing the type of interpretation of his (7) which Reichenbach wants to produce, but only on condition that it is not mixed up with token-quoting which would prevent it from doing (i) and (iii).

## VI. REFERRING TO SOMETHING AND REFERRING SOMETHING TO SOMETHING

So far I have only disposed of attempts to reduce the three presuppositions of language to the one presupposition that there is a subject-matter. It is almost as serious a mistake to try to reduce the remaining two to one by ignoring the speaker-listener or writer-reader polarity of the act of communication, even where we have succeeded in distinguishing between the act and its subject-matter. This can be shown more clearly by relying in what follows solely upon the distinction between act and subject-matter until the need for further refinement becomes obvious.

It has, I think, become evident that the demonstrative "this," for instance, doesn't normally refer to itself. It refers to something answering the description offered by noun-expressions such as "table," "paper," etc., which either follow or are understood. Only in special circumstances does it answer to that description itself (*viz.*, where the descriptive word is e.g. "token"), and only in still more special circumstances does it refer to itself in answering that description. We might then try to describe its essential function as follows. We might say that it identifies the object which answers the given or understood description by referring it in terms of the relation of spatio-temporal proximity (or sometimes merely the relation of having recently been mentioned) to the token within which the description is given or understood, *without necessarily referring to that token.*

For example, we might say that in its essential capacity Reichenbach's token (7)

- C. (i) refers to something answering the description "piece of paper" (dyadic referring)
- (ii) refers to something answering the description "token" (dyadic referring)

<sup>13</sup> But is more likely to be too fair, for a few lines below he goes on to say that the meaning of the operation symbolized by "this token" can only be formulated in a metalanguage. This is a distinct and alternative account which I shall discuss in section 8 under the heading of "systematic elusiveness."

- (iii) *identifies* the second by referring it to token (7) itself by means of the relation of spatio-temporal proximity (triadic referring).

If in addition to fulfilling these three functions it should happen to refer to its *own* token that will be due entirely to special contextual conditions which indicate that it is being put to alienated use—when, for example, there are no other tokens in the vicinity to which it could reasonably be interpreted as referring.

The negative proof of this interpretation consists in the collapse of Reichenbach's alternative interpretation. The positive proof consists in the context-variable reference of "this" (which supports both interpretations alike up to a point), and the fact that we don't normally<sup>14</sup> regard a statement which incorporates a personal pronoun or demonstrative as creating one of its own truth-conditions (which supports the above interpretation alone). If A says "I am overweight" he doesn't by so doing contribute anything to that statement's truth-conditions. Hence the logic of token-reflexives must presuppose a distinction between an expression-token and its subject-matter. For this reason too the reflexivity postulated by the above account is of an entirely innocuous kind. There is no question of the statement spinning its subject-matter out of its own entrails.

We can now go on to account for the use of "I," "here," "now," and their correlatives along similar lines. "I" would seem to identify the person it refers to by referring that person *inter alia* to its own token by means of the relation "utterer of." Similarly "here" would locate the object or place it refers to by referring it *inter alia* to its own token by means of the relation of simultaneity.<sup>15</sup>

Of all this group "this" alone would appear to have the accidental property of sometimes referring to its own token as well as that of fulfilling its

essential function of referring something to its own token. Because it is adjectival<sup>16</sup> in form it (along with "that") can qualify nouns such as "token" which describe its tokens. Hence some logicians have come to confuse its triadic referring function with its dyadic referring function and to imagine that any other expression with a triadic referring function must be reducible in some way to the operation of a pointing gesture or its verbal auxiliary "this." Even such alternative labels as "indexical sign" or "index" (Peirce)<sup>17</sup> and "indicator" (Goodman) appear to suggest the mistake which "token-reflexive" makes explicit.

To place the oddity of adjectives like "this" in correct perspective I offer in replacement a choice between "referential identifier" and "referential origin." The purpose of either is: (a) to mark the fact that "this" is but one among triadic referring expressions; (b) to distinguish their unique function, which is that of identifying, from whatever dyadic referring functions they may also possess; and (c) to indicate that they are the origins<sup>18</sup> of the different spatio-temporal systems of coordinates individually peculiar to each one of us, in terms of which we identify the bearers of our referring expressions.

It might also seem appropriate at this point to try to extend the same sort of analysis to the tense-inflections of verbs. Instead, I shall apply an amended version of this analysis to tense-inflections in the next section—for now it has become more urgent to attend to a major flaw. As it stands the analysis is quite unintelligible. How on earth can a statement refer one thing to another in terms of some relation without referring to both things and the relation between them? One might say that the object to which the statement refers an object, and the relation by means of which it does so, is contextually implicit. But this reply suffers from vagueness. How can it be made explicit within a statement that its physical relation to something

<sup>14</sup> When used non-tautologically "I am here" might be an exception, since once uttered it cannot be false, unless perhaps it has been telecasted by ventriloquism, radio, etc.

<sup>15</sup> But the relation of place or moment to token is sometimes merely that of having recently been mentioned.

<sup>16</sup> It has been brought to my attention by this paper's referee that "now" has an adjectival cousin, viz., "present," which can function in a similar way: cf. also "I" and "my."

<sup>17</sup> As has been pointed out to me by R. J. Butler, this allusion does scant justice to the fact, of which I had been unaware, that Peirce's account of indices is in terms of a triadic relation basically similar to triadic referring as outlined above (see *Collected Papers* [Cambridge, Mass., Harvard University Press, 1960], 2. 274 ff. and also A. W. Burks, "Icon, Index, and Symbol," *Philosophy and Phenomenological Research*, vol. 9 [1948-9]).

<sup>18</sup> Cf. Russell, *Human Knowledge* (London, Allen & Unwin, 1948), pp. 89-94. Yehoshua Bar-Hillel ("Indexical Expressions," *Mind*, vol. 63 (1954), pp. 373-374) argues that referential identifiers are indispensable to the *learning* of a language or adoption of a space-time frame of reference alone, and not to its use as Russell believes. But if referring expressions denote only because we denote something by them it seems to me that their use must always be anchored more or less directly and flexibly to some referential origin.

within its context (e.g., its token) is intended rather than accidental, if the statement doesn't actually refer to the object within the context and in that relation?

These difficulties would seem to be the direct result of treating statements and their components as detachable either from the people who make or the people who understand them. I have been discussing far too airily what an expression such as "I" does, or the function it performs, without specifying for whom it does or performs whatever it does or performs. This sort of negligence is both convenient and harmless where an expression performs much the same basic sort of function both for those who use and for those who understand it. But when we turn to the analysis of expressions which signify the individual difference between the participants in communication we must be wary.

The crucial distinction between those who participate in communication falls between whoever is the agent at the moment of utterance and the others, for to make a statement, ask a question, etc., is to act. Since this distinction is a condition of the speech-act it is presupposed by the content of all speech-acts, e.g., by all statements. Now we have already seen that when *A* says "I am overweight" he cannot possibly be using "I" to mean anything like "the utterer of this token." Being its agent he cannot refer to the token (which is an act) without changing the character of his statement from a report into something which creates its own truth conditions at least in part. The same restriction, however, does not apply to *B* considered as the respondent. He can refer to *A*'s act or to *A* as "the utterer of this token" without creating any of the truth conditions of his statement for the phrase "this token" is here used non-reflexively by *B* to refer to *A*'s token.

This does not amount to saying that "I am overweight" as uttered by *A* means "The utterer of this (that) token is overweight" as uttered by his listener *B*: for the subject-matter of *B*'s utterance would then include more than the subject-matter of *A*'s utterance. As we all know the utterance in *B*'s mouth which would be equivalent to *A*'s is either "You are overweight" or "He is overweight" or "*A* is overweight" respectively, according to whether he is addressing *A* or somebody whose attention is directed to *A* or somebody whose attention cannot be directed to *A* by any means so effective as the use of *A*'s name.

By such means as these the additional element

to the subject-matter of *B*'s utterance is avoided. Instead, where *B* uses "You" or "He" he refers to the same subject-matter as *A*, viz., *A*, by a comparable device: e.g., when using "You," he identifies his subject-matter, i.e., *A*, by referring it to his own token, i.e., as the person addressed by that token, without actually referring to that token. The equivalence between *A*'s and *B*'s utterances here depends upon the fact that at the moment of utterance the subject-matter about which *A* could possibly make statements partially includes and partially excludes the subject-matter about which *B* could make statements, and likewise the subject-matter about which *B* could make statements partially includes and partially excludes the subject-matter about which *A* could make statements. Each intentionally provides the other with that part of the latter's subject-matter about which he, the former, has to be silent, in order that both may focus upon a common subject-matter.

## VII. TENSE-INFLECTIONS

If tense-inflections and their colloquial equivalents are rightly classifiable as referential identifiers, and I think they are, then it likewise follows that the Reichenbachian and highly prevalent analysis of such expressions is quite wrong. Minus its various refinements, which we can here safely ignore, this analysis would translate tense-inflected statements into statements which characterize the event described as before, after, or simultaneous with the token of the statement without using any tense-inflected or colloquial equivalent. Thus "Brutus killed Caesar" is taken to mean "Brutus kills (timeless present) Caesar and this killing takes place (timeless present) before this token" or less concretely "Brutus kills (timeless present) Caesar at time *t* and *t* is before this token."

An attempt of this sort to translate tense-inflected verbs into non-inflected verbs rests simply upon a confusion between dyadic and triadic referring. If a tense-inflection did refer in the dyadic way *inter alia* to its own token, the translation which exhibited this would certainly have to eliminate tense-inflections to escape infinite regress. But it cannot refer in this way to its own token for the same reasons that prevent referential identifiers in general from referring dyadically to their own tokens. The most that a speaker *A* does for his token in stating "Brutus killed Caesar" is to offer that token to his listener

*B* for subject-matter, so that *B* can say "Brutus killed Caesar before *A* made his statement." Here, however, *B*'s subject-matter is in excess of *A*'s. If on the other hand *B* had wished merely to corroborate *A*'s statement all he would have needed to do would have been simply to repeat it. Both *A*'s and *B*'s statement would then have exactly the same subject matter. This is because both *A* and *B* offer their own tokens and accept the other's for purpose of identification without referring (dyadically) to either.

We have a comprehensive solution here to a problem which was first conceived in a rather arbitrarily restricted form.<sup>19</sup> If we translate the force of the inflection in "Brutus killed Caesar" by stating that the killing takes (timeless present) place before the date at which the statement is made then no two statements of this same sentence-type can be synonymous unless they are uttered simultaneously, since each would refer to a different date or occasion.<sup>20</sup> The same difficulty, however, would affect the synonymy of some soothsayer's prediction "Brutus will kill Caesar" uttered before the event and a report, say, by Plutarch of the same incident uttered after it had happened.<sup>21</sup> And we now can see that this problem in its turn is nothing but a special version of the more general problem<sup>22</sup> created by the misunderstanding of other referential identifiers besides those of tense.

Referential identifiers in general fall into groups of antithetical terms such as "I-you-he-she," "We-you-they," "here-there," "now-then," "this-that," "past-present-future," etc. Each member of any such group can be used in a sentence which makes exactly the same statement or asks the same question or in some cases gives the same command, etc., as another sentence containing one of its correlative opposites within the same group. In all cases the solution is to be found in the distinction between dyadic and triadic referring in the context of the potentially reciprocal speaker-listener (writer-reader) act of communication.

<sup>19</sup> A. Duncan-Jones, "Fugitive Propositions," *Analysis*, vol. 10 (1949).

<sup>20</sup> Jonathan Cohen's valiant effort to cope in Reichenbachian terms with this difficulty in "Tense Usage and Propositions," *Analysis*, vol. 11 (1951), and "Postscript (1954)," *Philosophy and Analysis*, ed. Margaret Macdonald (Oxford, 1954), p. 188, is vulnerable to all the criticism outlined in sections 1-5 above: but cf. § 19 of his *The Diversity of Meaning* (London, 1962) for a rather different approach.

<sup>21</sup> One could no doubt dispense with tense-inflections by means, for instance, of a combination of "before" or "after" and "now." Such a language would, however, be less sophisticated than ours just because non-simultaneous predictions or reports of the same event would be imperfectly synonymous, since "now" in each case would refer to something different.

<sup>22</sup> Bar-Hillel, *loc. cit.*, pp. 364 ff., disposes of the general problem in terms of a different sort of triadic relation from mine. An "indexical" sentence paired with a "pragmatic context" (which is a non-linguistic event) may form a "judgment" which "refers pragmatically" to the same proposition as differently constituted judgments. Besides leaving the key concept of pragmatic context rather vague, as he himself admits (p. 371), this solution seems to distend the notion of dyadic referring beyond recognition.

#### VIII. SYSTEMATIC DISCREPANCY, INCONGRUENCE, AND ELUSIVENESS

The obsession with the pointing gesture as a model for referential identification in general achieves perhaps its most spectacular variant in Ryle's account of what he calls "index words" (*The Concept of Mind*, pp. 186-98). He describes (p. 197) identifiers as systematically elusive in the following sort of way:

I can point to any other thing with my index-finger, and other people can point at this finger. But it cannot be the object at which it itself is pointing. Nor can a missile be its own target, though anything else may be thrown at it.

This would seem to bear a stronger moral than the rather similar illustration of Wittgenstein's which I took as my starting point, for Ryle's conclusion is that "To concern oneself about oneself in any way, theoretical or practical, is to perform a higher order act, just as it is to concern oneself about anybody else" (p. 195).

If Ryle claims to be concerned with the reference of activities, self-directed or otherwise, then the way in which he has illustrated their systematic elusiveness conceals a confusion between an activity and a condition. The condition of being directed at something is not *ipso facto* an activity. Index-fingers, weather-vanes, magnetic needles, and missiles are directed at something in the physical sense that the thing lies on the projection from the sharpest end of their longitudinal axes or, in the case of missiles, on any part of their untraversed trajectory. This is a physical condition, and not an activity, and it is logically implicit in a condition described in this way that any object in this condition cannot be directed at itself—so that in this way such an object is systematically elusive.

Besides the distinction between the physical conditions and directive activities, further distinc-

tions among the latter should be made between directing something at something, or putting something in the condition of being directed at something (e.g., aiming a gun), and directing somebody's attention to something by conventional or other devices. To ignore both sets of distinctions would be to assimilate the extremes, i.e., the physical condition and directing somebody's attention. But just this may happen if we miss the fact that such descriptions as "index finger," "missile," "weather-vane," "magnetic needle," etc., are theory-laden terms.

Once we describe an object in one of these ways it may seem reasonable to say both that it *does* something, i.e., points, and in some cases that it actually directs our attention to something, and in a conventional way, while being in the physical condition of being directed at something. But the sense in which it points is more than the sense in which it is in the physical condition of being directed at, only because it is an instrument which we use for directing something at something, and in some cases for directing peoples' attention to something. Strictly speaking the object does not do the pointing but is instrumental to our engaging in activities of the sort outlined.

Now obviously the activity of directing somebody's attention to something by conventional means does not escape itself in the way that an index finger escapes itself. If any further argument is necessary, an extension of Ryle's missile analogy should be enough. When one casts a clay pigeon in a certain direction one certainly is not casting it at itself, and in that sense what one does is systematically elusive. But, thanks to the conventional significance of the object one is inviting others to shoot at it. One is directing their attention to it, and expressly so—not inadvertently, unless it was through sheer inadvertence that one selected it as a missile rather than some other object.

Similarly in the case of personal pronouns, tense-inflections and other expressions besides "this" or "that." When the author of any such utterance is engaged in the activity of communication what he is doing doesn't escape his attention in the act of doing it. It is not systematically elusive in this sense then. It would be more accurate to say that what he is doing is systematically incongruent with what he is referring to, though it is not systematically incongruent with what the respondent may refer to. On the other

hand what the respondent may refer to is systematically incongruent with what the respondent is doing in so referring. Communication does not consist, as Ryle would appear to think, of a game of leapfrog between higher order activities, viz., a game of one-upmanship between metalanguages; for language is communication. Come to think of it, a game of leapfrog itself isn't a series of meta-games but just the one game.

The reason why self-concerning activities on the whole may not be the subject of communication consists mainly of a limitation of language or thought directed specifically, though in a rather wide sense, to theoretical purposes. Theoretical activities except in limiting cases are indeed systematically elusive, for an objectively true statement cannot supply any of its own truth-conditions. But practical activities are not limited in this way.

When the speaker communicates, that part of the object of his attention which is incongruent with what he does refer to in speech is occupied by his activity of communicating. And the reason why it is incongruent is because the act of attention is inseparable from this part of its activity. It is in fact this interpenetration of attention and object of attention that constitutes an intentional action as intentional. Perhaps this can be more clearly seen in the case of activities other than acts of communication. If I decide in a given set of circumstances to water my garden on Saturday my awareness of what I am about to do must be regarded both by myself and others as essential to my being about to do it provided I do have a genuine choice in the matter.<sup>23</sup>

In cases of this sort one can then proceed in a quite separate activity to communicate one's decision by expressions of purpose such as "I am going to water the garden on Saturday." But this possibility of communication doesn't really remove the systematic incongruence between what I and others may refer to with respect to my decision. In fact the incongruity acquires an extreme form which I shall call "systematic discrepancy."

The decision expressed by "I am going to water the garden on Saturday" is, for instance, based upon such considerations as the speaker's having a hose, a plentiful supply of water, and no conflicting engagements. On the other hand, the listener's prediction "He is going to water the garden on Saturday" is open to these same considerations

<sup>23</sup> For a fuller and more guarded analysis of such expressions along these lines see the author's *Choice and Chance* (Oxford, 1961), chaps. 6-7.

plus the consideration that the agent has decided or will decide to water the garden. To the listener all these factors are externally related. The presence of the rest need not be affected by the absence of one. But to the speaker at the moment of his decision only the first three types of factor are external to each other, for they are internal to or constitutive of his decision.

#### IX. THE MIND-BODY ISSUE

Ryle employs the notion of self-elusiveness to demonstrate that "I" is not the self-appropriated soubriquet of each ghost-in-the-machine. He is concerned lest we infer from the supposed fact that each of our self-directed activities eludes itself that the "I" refers not merely to the person whose activity it is, but to something imperceptible which haunts these activities. This no doubt is a very proper concern, but one which would be better served by noticing that the supposed fact isn't a fact at all.

If the logic of personal pronouns is absolutely central to the mind-body issue as I suspect it may turn out to be, then it would seem that we cannot offer any of the classic solutions to this issue. Neither the monistic types of solution such as materialism, behaviorism, physicalism, idealism,

or neutral monism, nor the dualistic ones such as epiphenomenalism, psycho-physical interactionism, or psycho-physical parallelism will do. They all appear to ignore the systematic incongruence presupposed by communication and hence, I assume, by any possible conceptual framework. They all would seem to attempt a reduction of two of the presuppositions of communication, viz., the existence of a speaker (writer) and the existence of a listener (reader), to a third, viz., the existence of a subject-matter.

Had systematic elusiveness proved to be of the weather-cock sort postulated by Ryle, then it could have been fitted into some sort of physicalistic frame, for it is just the sort possessed by index fingers, etc. On my interpretation, however, systematic elusiveness is simply a device for confining the subject-matter to the relatively limited though large area which remains unaffected by the structural discrepancies between the speaker's and the listener's point of view. These discrepancies must fault any attempt to stratify the world monistically or dualistically.

But unless or until the logic of personal pronouns can be shown to be absolutely central to the distinction, say, between my tooth and my toothache, this must at best remain a highly speculative conclusion.

## VI. THE PROBLEM OF EVIL AND THE PROBLEM OF SUFFERING

G. SCHLESINGER

### I

THE world is full of suffering; God is either helpless to prevent it, in which case He is not all-powerful, or does not choose to prevent it, in which case He is not all-good. It is not surprising that even after centuries of debate some should regard this as the most effective argument against theism, while others think of it as wholly ill conceived. Regardless of a philosopher's convictions, when considering a problem such as this, which touches upon vital issues like Divine omnipotence and providence, man's fate and freedom, he naturally finds it hard to keep to cool analysis untainted by his hopes and fears. However, even on the purely logical plane the problem has its unique difficulties, because it involves the highly elusive concept of infinity as applied to power and benevolence. I shall want to suggest that, in addition, the concept of infinity is involved, in a hitherto un contemplated way, in the question of whether there is ineradicable evil, and that therein lies the key to the dissolution of the problem. But first I must clarify a number of points, and I propose to do so by considering very briefly some of the well-known arguments that have been employed in the context of our topic.

### II

To begin, let us agree that the theist is not entitled to say: "I have no proof (because of my limited understanding) that omnipotence, benevolence, and the existence of evil are mutually compatible, but I have faith that they are, just as I have no proof that God exists but I have faith that He does." For it is not at all "just as." The "leap of faith" required in the former case is very different from that of the second, and this should not be overlooked. Lack of a proof that God exists can at most provide grounds for agnosticism; lack of a proof that the problem of evil has a solution, on the other hand, provides *prima facie* support for atheism. But nevertheless, if the theist

merely says, "I have no solution for the problem of evil but I have faith it has a solution," then he commits no fallacy; he only puts himself into a logically weak position which a genuine suggestion for the solution of the problem could avoid.

The situation may be viewed as even worse with regard to the argument which maintains that Divine goodness is entirely different in kind from human goodness and should consequently not be judged in terms of notions formed on the basis of our acquaintance with the latter. Thus the fact that moral and physical evil seem to exist and to be condoned by Him should not be taken as a sign of deficiency in His goodness because His ways are inscrutable to limited human minds.

The obvious objection that may be raised against this is that we have no other notions of good and bad except those appertaining to human situations, where, for example, the condoning of unnecessary suffering is always condemned. Therefore, we must either be allowed—even when talking about God—to use the only moral notions we have, or we cannot say anything meaningful on the subject. (There seems to be good ground for the atheist's complaint that if the theist does not choose the first alternative his position amounts to saying, "I do not know what I mean when I maintain that no evil can be attributed to God but I have faith that it means something significant," which is quite absurd. The atheist, on the other hand, can plausibly lay claim to a clear notion of what is morally wrong, namely, to cause or condone unnecessary human suffering, and thus has a notion of what it is to attribute such wrongs to God.)

A variation on the previous argument is the one which holds that it is wrong to say that evil is permitted to exist in the world, since, simply by definition, evil is that of which God does not approve and does not permit to exist. It seems obvious that, unless we have independent means of judging whether or not all of God's actions are good, the theist's claim that no moral imperfection can be attributed to Him is empty verbiage.

## III

Some proposed solutions of the problem attempt to diminish or eliminate evil. At times it has been argued that there is far less suffering in the world than one would unreflectingly assume. Some have eloquently explained that pain is not additive and that there is thus no such thing as a "suffering humanity" transcending the individuals it consists of. Others exhort us (1) not to take for granted all the joy and moral beauty there is in the world which by far outstrips the amount of existing pain and moral ugliness, or argue (2) that all suffering ultimately brings about a greater good and should be viewed always as a preliminary to the latter.

It suffices to say that, while there is much wisdom in many of these arguments, they do not mitigate logically the problem we are facing. Multiply or divide the amount of pain in the world by a billion, and its incompatibility with Divine goodness and omnipotence is not affected. Were it that merely a single individual had to endure unnecessarily a slight inconvenience for a brief moment, the problem—even though it would not "hurt" so much—would remain just as if everybody caused and endured suffering continually.

Another argument claims that there is no real, ineradicable evil because all suffering can be avoided. Sometimes it is claimed that the human soul has infinite resources and it is in the power of every individual to transcend pain and feel happy irrespective of his circumstances. It is not our concern here to examine the various assumptions involved in this argument—let it be conceded that each one of us has it in him to reach the peak of disinterested spirituality. It is still a fact that many people are unhappy, apparently leaving their spiritual resources untapped out of ignorance or negligence. The skeptic may still pose the problem: Why are spiritual potentials allowed to go unrealized; why, even if there is no need, is it nevertheless possible to suffer?

Similarly, when it is said that suffering is always chastisement for transgressions, the skeptic may ask, apart from everything else, why have opportunities to sin been created? It may be the case that the reward for righteousness is incomparably greater than the punishment for wickedness, and that this reward tastes a thousand times sweeter now that there exists the possibility of sinning and being punished. However, the skeptic will go on urging, why not let human nature be different and

man be made so as to be able to enjoy any amount of pleasure without needing a contrast?

## IV

Another famous line of argument is to concede that evil exists, indeed necessarily so. If we had access to God's whole plan—it has been said—we might be able to appreciate that a given region of human experience, which, though in isolation looks ugly, viewed as a part of the entire pattern, enhances the over-all beauty and goodness of the whole. But even if the universe is much better because of having some evil in it and the same result could not otherwise be achieved, this is only so in the context of the existing laws of nature. Why did God not make nature's laws such that the amount of the advantages we enjoy should remain exactly the same as at present but there be no need to pay for it in pain?

There have also been many attempts to show that the necessity for the existence of evil is of the logical kind. Every person is the sum total of his experiences, it has been said, and it is therefore logically impossible that any given individual suffer less pain and anguish and enjoy more peace and happiness; for alter a man's experiences and you are no longer dealing with the same person. A more common and perhaps more serious attempt is based on the contention that a desirable quality, like the freedom of the will, for example, presupposes moral and physical evil (since by definition man's will is not free unless he can choose to do harm and creatures vulnerable to harm existed). It has also been maintained that there are other precious qualities like heroism which is analytically connected with danger, charity with want, compassion with pain, and hence it is logically impossible to have these desirable human traits without their unwanted prerequisites. We shall assume the reader is acquainted with arguments that have been produced to defeat these moves. For our purposes it does not matter if we also assume these arguments to be wholly effective. But we must note that in general they all lead to the conclusion that evil is only contingently but never logically necessary.

## V

It has already been pointed out that if the world had been different and the amount of suffering in it had been a minute fraction only of what it is

at present, this would have made great emotional, but no logical difference. It would seem that there exists a certain minimal level such that if the amount of suffering were reduced below it the problem of evil would not have occurred to anyone. Logically it would have been there just the same.

It should be added that even if no one ever experienced any pain or discomfort but some were deprived of positive happiness the problem would remain.

Taking once again our moral criterion from human affairs, if I have the opportunity to cause a person extreme happiness without harming anyone, without any expense or effort on my part, and I am aware of this, (even though *he* may not be) then if I refrain from doing so, I am morally reprehensible.

Thus, alter the world; diminish the amount of suffering until it vanishes altogether, and the problem remains: Why did not God make men much happier than they are?

Even the most satisfied man on earth could make up an endless list of requirements the fulfillment of which would increase his happiness. Thus it would not help if every individual was as happy as the happiest man. And even if everyone enjoyed such a state of supreme satisfaction, no longer asking for anything and claiming to have no unfulfilled desires, this would not show that God has discharged His "moral obligations" toward His creatures. For example, whether my dealings with *X* are beyond moral blame is not solely determined by the fact that *X* is completely satisfied. If *X* is a musically gifted child in my care whom I have provided with all his physical and emotional needs but whose musical education I have neglected, thus depriving him of a mode of enjoyment the possibility of which is wholly unknown to him, I am nevertheless blameworthy. Similarly a deity who is aware of the much greater possibilities for happiness that *could* be the lot of His creatures, but does not realize them, seems to be blameworthy irrespective of what these creatures think and feel.

There being no *prima facie* case for saying that the greater possibilities for happiness are finite, God's inability to create the greatest state of happiness is seen to be no different from his inability to create the greatest integer. Neither diminish His might, if we agree that He need not accomplish what is logically impossible. We cannot ask that He go against the laws of logic for we do not know what it is we are asking of

Him; and when He does not do what is logically impossible, there just is no feat which we can say He failed to accomplish. (His inability to do what is logically impossible of course does not impair His omnipotence.)

What we have to conclude, therefore, is that the amount of pain and joy present in the world is entirely irrelevant and cannot be introduced as evidence concerning the moral nature of God. (There is of course scope left for investigating whether the question "Who (when and how), is happy or suffers?" may not provide a clue to the nature of Divine benevolence; but the mere fact of the presence of pain or joy in any amount points to nothing.)

The problem of evil arose originally because the proposition "God is omnipotent and no evil can be attributed to Him" and "There is a state of affairs in the world which a perfectly moral Being should not tolerate" are inconsistent. In order to do damage to the theist's position it was essential that the last proposition be thought of as based on observation. Evil has been observed in the world and this speaks against the belief in Divine goodness and omnipotence. In posing the problem, we made the implied assumption that the universe could be different from what it is now and then the problem would disappear. But now we have seen that—alter the universe as you wish—it does not affect the problem. It therefore does not arise in the first place.

The problem of evil disappears, then, because we have been shown that while the principle "It is morally wrong, when it does not interfere with the welfare of others, not to do as much as one *possibly can* to make others happy" may apply to humans, it *logically* cannot apply to an Omnipotent Being. (This of course does not amount to one of the attempts, considered briefly earlier, to solve the problem by suggesting there is an ineffable kind of evil, and it is of *this* that we are speaking when maintaining that no evil can be attributed to God. For here we are talking about a familiar notion of evil and show that for reasons perfectly unmysterious it cannot be present in the actions of God.)

We have, of course, provided absolutely no explanation why the world is not a much better place, and one may find this exceedingly frustrating. But no matter how much one may yearn for an explanation, the lack of it does not warrant a confusion between the problem of evil and the problem of suffering. True enough, if the theist is

right there are many pertinent questions we may pose: Why is not everyone at least free from positive suffering? Or why is not everyone at least in a state of minimal happiness all the time (given a reasonable definition of "minimal" happiness)?

But the possibility of posing these questions must not be construed as a possibility of questioning the theist's contention that evil is not to be attributed to God on grounds of the experienced character of the world we live in.

*University of Pittsburgh*  
*and*  
*Australian National University*

# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson  
Kurt Baier  
Richard B. Brandt  
Roderick M. Chisholm  
L. Jonathan Cohen  
James Collins  
Peter Thomas Geach  
Adolf Grünbaum

Norwood Russell Hanson  
Carl G. Hempel  
Jaakko Hintikka  
Jean Hyppolite  
Raymond Klibansky  
Helmut Kuhn  
Benson Mates  
John A. Passmore  
Günther Patzig

Oliver L. Reiser  
George A. Schrader  
Wilfrid Sellars  
J. J. C. Smart  
Wolfgang Stegmüller  
John Wild  
G. H. von Wright  
John W. Yolton

---

VOLUME 1/NUMBER 4

OCTOBER 1964

## CONTENTS

- |   |     |  |     |
|---|-----|--|-----|
| I. HENRY E. KYBURG, JR.: <i>Recent Work in Inductive Logic</i> . . . . .    | 249 | III. PETER WINCH: <i>Understanding a Primitive Society</i> . . . . .           | 307 |
| II. GERHARD GENTZEN: <i>Investigations into Logical Deduction</i> . . . . . | 288 | IV. HERBERT SPIEGELBERG: <i>Toward a Phenomenology of Experience</i> . . . . . | 325 |
|   |     | V. <i>Corrigenda</i> . . . . .   | 333 |
-

# AMERICAN PHILOSOPHICAL QUARTERLY

## POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with a wide margin, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively—also being typed with wide margins and spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## REPRINTS

Authors who are subscribers will receive 50 reprints gratis. Additional reprints can be purchased through arrangements made when checking proof.

## SUBSCRIPTIONS

PLEASE NOTE: Beginning with the year 1965 the price *per annum* for all individual subscribers will be six dollars and the price *per annum* for institutions will be ten dollars. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues will continue to be sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscriptions and back orders may be addressed directly to the publisher (University of Pittsburgh Press, University of Pittsburgh, Pittsburgh, Pennsylvania 15213).

\* \* \*

The *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.

Second-class postage paid at Pittsburgh, Pennsylvania.



## I. RECENT WORK IN INDUCTIVE LOGIC

HENRY E. KYBURG, JR.

THERE has been a great deal written about inductive logic in recent years. For the purposes of this article I have taken the year 1951 as a benchmark; what transpired before that I take to have been largely assimilated by the philosophical world. Thus on the lower side of the benchmark we find Carnap's monumental *Logical Foundations of Probability*, Russell's *Human Knowledge*, Kneale's *Probability and Induction*, Williams' *The Ground of Induction*, Reichenbach's *Theory of Probability*, and von Wright's *Treatise on Probability and Induction*. Taking for granted some familiarity with these materials (but only familiarity of the most general sort) I propose to deal with the developments of 1951 and later.

It is difficult to know just what should be regarded as inductive logic. There are writers (K. Popper, for example) for whom there is no such thing as inductive logic, and writers (S. Toulmin, for example) for whom all interesting logic is more or less inductive. There are writers who feel that it makes sense to talk about inductive *logic* only after induction has been justified (or vindicated) in general (A. W. Burks), and there are writers who feel that inductive logic is nothing but a theory of probability (R. Carnap), and there are those who think that probability has nothing to do with the matter (Popper). There are those who think the question of causality is fundamental (A. C. Ewing), and those who think that the fundamental questions concern decisions in the face of uncertainty (C. W. Churchman). I have felt that I had to deal with all of these issues, to some degree; but I have tried to stray no further from the central topic of this essay, inductive logic, than my conscience forced me to stray.

### I.

One question about inductive logic concerns whether or not inductive conclusions are logically necessary.

There are two ways in which necessity may be attributed to inductive conclusions: The conclusions themselves may be regarded as certain, i.e., as following necessarily and incorrigibly from the evidence for them, and the conclusions themselves may be taken to express some sort of necessity—such as causal necessity. Ideally, Mill's methods were to reveal necessary connections in a necessary manner, and thus to involve necessity in both senses.

There are few philosophers left who are still willing to maintain that scientific method yields certainties, but there are some. An unusually extreme version of the view that induction yields necessity is argued by W. I. Matson,<sup>1</sup> who maintains a traditional rationalist point of view, according to which there is no distinction between the necessary and the contingent—all facts are necessary—and there is no such thing as induction (we do not generalize from experience) and hence no *problem* of induction. Matson claims that we achieve mathematical necessity in science, though he does not say how. A counter-instance "would not be a change of the course of nature, but an annihilation of nature."<sup>2</sup> It is enough to refute this view to point out that our scientific expectations are not always fulfilled; nature is repeatedly, to use Matson's term, "annihilated." In seeking an inductive logic or a scientific method, we are seeking a way of evaluating the likelihood of such an annihilation.

A more plausible way of achieving certainty for inductive conclusions is provided by ordinary-discourse analysts. Thus J. O. Nelson<sup>3</sup> writes, "... scientists all the time conclusively establish ... hypotheses." For example, we may speculate about the existence of life on Mars; when scientists have amassed enough evidence of the right sort, they do not say that it is "highly probable" that there is life on Mars, but that it is now certain, it has now been conclusively established, that there is life on Mars, because "It has been conclusively

<sup>1</sup> Matson, "Against Induction and Empiricism."

<sup>2</sup> *Ibid.*, p. 157.

<sup>3</sup> Nelson, "The Confirmation of Hypotheses."

established by spectrographs that the green spots are lichen." To talk in any other way, according to Nelson, is to abuse language. But it is clear that the issue here can no more be settled by taking a poll of common usage than can the issue of whether a whale is to be regarded as a fish or a mammal. As L. Resnick points out<sup>4</sup> in his comments on Nelson's article, "probability" is used in philosophy, as in science, as a technical term.

In another article<sup>5</sup> Nelson tries an ingenious device for achieving certainty for inductive conclusions: that of leaving the quantifier off the inductive conclusion. Thus he claims that we do not argue from "some crows are black" to "all crows are black," but merely from "this crow is black" and "that crow is black" to "crows are black." This inductive conclusion is, in a sense, *conclusively* verified by the earlier observations. "... it is precisely because their ordinary mode of expression is characterized by ambiguity that inductive generalizations can possess (conclusive) verification and evidence."<sup>6</sup> If it were not for this ambiguity, we could never have evidence relevant to future actions and commitments, for in the future we are concerned with unobserved crows; we have never observed one of them, and thus we have no evidence concerning them.

Again Nelson quickly finds an antagonist, J. E. Llewellyn,<sup>7</sup> who accuses him of abusing technical language; Llewellyn says that "inductive" is a logicians' word, and they ought to be allowed to set its usage. "Crows are black," is not a proposition at all—for when some crows are black and some are not, it is true and false at the same time. The very least a logician has a right to demand is that statements of science lack this kind of ambiguity.

W. Kneale, it will be recalled, presented a doctrine according to which general synthetic propositions of a certain sort could be considered certain, although they could not be known *a priori*.<sup>8</sup> J. P. Day, who follows Kneale's lead in other ways, follows him in this as well; he argues that not all ampliative induction is probable, "For we do often call generalizations true or certain."<sup>9</sup>

I think it can more plausibly be maintained that

certainty in this sense is a liability to scientific conclusions than that it is an asset. As long as we can regard a counter-instance as possible the conclusion is not certain; and to regard a counter-instance as *impossible* is just to commit oneself to explaining it away, come what may—i.e., it is to regard the conclusion as analytic and uninformative.

These writers do call our attention to something important, however: The idea of practical certainty as a level of confirmation at which we can detach scientific conclusions from the data that support them. It should also be noticed, however, that practical certainty is a flexible concept: In Nelson's example, were one to fix the astronomer with a beady philosophical eye and say, "Are you absolutely certain that there is life on Mars? Has no one ever misread a spectrograph?" he would of course say, "Well, it is at any rate overwhelmingly *probable* that there is life on Mars."

It is also true that many of the arguments that are used in science—arguments that for the moment I am lumping together as inductive, perhaps in defiance of etymology, are conclusive, or at least give the appearance of being conclusive, in the sense that the conclusion follows deductively from premisses—general as well as particular—that we *accept* in a given inquiry. The analysis of these arguments is the analysis of *demonstrative induction*. The most extensive piece of work on demonstrative induction is that of G. H. von Wright.<sup>10</sup> In this book he is mainly concerned with the logical problem of analyzing the inferential mechanism of induction rather than either the psychological problem of discovery or the specifically philosophical problem of justification. Chapters Four, Five, and Six are on the logic of *necessary, sufficient, necessary and sufficient* conditions. (The other chapters are concerned with non-demonstrative induction, but that is considered from the same point of view as consisting essentially in the *probable* eliminations of causal laws.) We have here an elaborated version of Mill's canons of induction.

In each case considered by von Wright, we begin with the knowledge that the necessary, or sufficient, or sufficient and necessary condition is to be found among a list of candidates which is generally

<sup>4</sup> Resnick, "Confirmation and Hypothesis."

<sup>5</sup> Nelson, "Are Inductive Generalizations Quantifiable?"

<sup>6</sup> *Ibid.*, p. 61.

<sup>7</sup> Llewellyn, "Unquantified Inductive Generalizations."

<sup>8</sup> Kneale, "Some Aspects of Probability and Induction," and *Probability and Induction*.

<sup>9</sup> Day, *Inductive Probability*, p. 256.

<sup>10</sup> von Wright, *A Treatise on Induction and Probability*.

assumed to be finite. It is easy to see that, given a list of possible conditions, one could construct a list of all the possible causal laws relating those conditioning properties to a given conditioned property. Even if all the possible combinations of causal factors in this list are allowed, von Wright shows, it is still possible for inductive elimination to lead to certainty: When all the possible sufficient conditions but one are eliminated, the one that remains will be the conjunction of all the possible conditioning properties; and when all the possible necessary conditions but one are eliminated it turns out to be the disjunction of all the conditioning properties.

It is obvious that this is not a very interesting result. By imposing special conditions on the form of the properties in which we are interested, we can get more interesting results; but in any event we can get no results which we could not get by composing a list of possible causal laws, and attempting to eliminate all but one. Indeed, if the interest of constructing an inductive logic is to throw light on the sort of thing scientists actually do, the latter procedure is far more realistic than the former. It is far more sensible to suppose that in a determinate inquiry we have narrowed the plausible possibilities down to three or four specific laws, and that we then proceed to eliminate them as best we can (consider, e.g., the well-known procedure of Pasteur), than it is to suppose that we write down a list of properties that *might*, in some combination or other, be conditioning properties, and that we then try to eliminate all but one of the much larger list of possible laws that we can construct on the basis of this list of possibly conditioning properties.

In principle, at any rate, there is nothing problematic about the logic of elimination: the law is a general statement, and the observation report can be a singular statement that contradicts the law. But something might well be written about the construction of the sets of laws of which we are to eliminate all but one; this might be a contribution to the actual practice of science as well as to the logic of induction. Where does a useful set of possible laws come from? On what grounds can we accept the hypothesis that one of these laws will be found to hold? Do such sets of hypotheses exhibit an internal structure? They clearly do not exhibit the complex and exhaustive structure that Mill and von Wright suppose they do. But so far

as I know nothing has been done along these lines.

A way of handling the elimination of an infinite number of hypotheses, where the inference concerns natural kinds, has been developed by myself.<sup>11</sup> Here is an example of a common form of scientific argument, which can be explicated by the techniques of that paper: A species *S* falls under a certain genus; a property *P* belongs to a certain class of properties. We know that all the species of that genus are such that if one member of the species has a given property belonging to that class of properties, then all members of that species have that property. The inductive argument then proceeds from this generalization and the observation of a particular instance of one of the species in question, to a general conclusion: All members of this species have that property. As a typical example, a pure chemical compound will have (under standard pressure) exactly one melting point. If one sample of compound *X* melts at *T* degrees centigrade, then all samples of that compound will melt at *T* degrees centigrade. We do not achieve certainty, of course, for our inductive conclusion that all samples of compound *X* will melt at *T* degrees under standard pressure. There are two sources of doubt: (1) the generalization concerning natural kinds is only probable—though there could be argument concerning where the element of uncertainty entered and how—and (2) the particular statement that serves as a premiss may be subject to doubt: we can only have *good grounds* for believing that the sample we tested actually did melt at *T* degrees, because we must allow for experimental error. The same two comments apply to von Wright's reconstruction, of course. In any real situation the hypothesis that one of a set of possible hypotheses holds is only probable, and in any interesting case the relevant "observations," like the observation of the melting point of a substance, will only be probable.

S. Nowak<sup>12</sup> carries the analysis of necessary and sufficient conditions into the field of statistics and derives some very interesting results. As an example, suppose that *A* and *S* are necessary and sufficient for *B*; if *S*<sub>1</sub> is an essential component of *S* (i.e., if *S* is composed of *S*<sub>1</sub> and *S*<sub>2</sub>), then the probability of *B* given *A*,  $P(B/A)$ , will be the product of the probability of *S*<sub>1</sub> given *A*,  $P(S_1/A)$ , and the probability of *B* given *A* and *S*<sub>1</sub>,  $P(B/AS_1)$ . If these probabilities are interpreted as relative frequencies, this relationship is empirically test-

<sup>11</sup> Kyburg, "Demonstrative Induction."

<sup>12</sup> Nowak, "Some Problems of Causal Interpretation of Statistical Relationships."

able, and may throw light on the possible causal relationship between  $B$  and  $S_1$ . Results like this, Nowak shows, are of considerable practical and theoretical importance in actual statistical work in the constructing and testing of statistical hypotheses. They enable us, for example, to discuss such things as spurious correlation and spurious independence more carefully and rigorously than we could without the analysis into necessary and sufficient conditions.

The attempt to find necessity, not *for* inductive conclusions, but *in* them is rather more successful. There are many philosophers who fear that unless such a necessity can be found, inductive inferences *will not* be justifiable. A. C. Ewing,<sup>13</sup> for example, fears that "if inductive inferences are never justifiable, science and superstition are in the same boat." He finds the required necessity in the notion of causality, and he presents as a prototype of a causal argument the following: there must (probably) be a causal connection between fire and heat, because when I put my hand near the fire it feels warmer. That this should happen by chance twenty times is  $(\frac{1}{2})^{20}$ . "At least we know that the antecedent probability of its feeling warmer on a particular occasion cannot in the absence of any relevant causal law be greater than one in two. Therefore the antecedent probability of its feeling warmer in the absence of a causal law on all of twenty occasions when I put it near the fire is already less than  $(\frac{1}{2})^{20}$  . . .".<sup>14</sup> And thus the probability of there being a causal law "seems to follow clearly."

It is difficult to deal with this sort of argument aside from pointing out that whatever happens is extraordinarily improbable, antecedently, however we interpret probability. H. Freudenthal<sup>15</sup> makes the same observation concerning this particular technique of probability inversion: what he observes at Monte Carlo during any short period of play is something that could happen "by chance" only once in a thousand years; but from the fact no profound conclusions follow. Furthermore, while we all agree that the conclusion that fire will warm my hand the next time is probable, no mechanism, no real *logic* is provided by the causal analysis which will enable us to distinguish between arguments whose conclusions do have this characteristic necessity, and arguments that do not.

The causal analysis does, however, point to a

characteristic of many inductive conclusions. They are regarded as formulating necessities, though not necessary themselves. Water not only *does* boil when heated to  $212^\circ$  under standard conditions, but it *must* boil then. One way of calling attention to the necessity embodied in such statements as "Water boils at  $212^\circ$  under standard pressure" is to point to the fact that these statements are used to support counterfactual assertions, such as, "If the substance in the beaker were water, it would have boiled at  $212^\circ$  under standard conditions." Such an assertion could be sensibly made, and would be regarded as true, even though the antecedent is false. A *merely* general statement, one which does not embody any notion of causality, cannot serve to support counterfactuals: "All the people in this room are men" does not support the statement "If Mary were in this room, she would be a man."

## II.

There are many writers of many persuasions who feel that inductive conclusions often express some kind of necessary connection, but there are few who make this necessity bear much of the burden of inductive logic. Most of those who do feel that the concept of necessary connection enters into the canons of inductive logic simply formulate postulates, as did Russell in *Human Knowledge*, which are alleged to be sufficient to establish (with the help of empirical evidence) the cogency of inductive conclusions. The postulates are not, like Ewing's postulate of causality, to be certified on metaphysical grounds, but neither are they to be regarded as ordinary well-established material hypotheses like von Wright's lists of possible conditioning properties or my general conditional statements about natural kinds.

The best known attempt to found both inductive logic and inductive justification on broad, factual assumptions is that of A. Burks.<sup>16</sup> Burks argues that we must turn to presuppositions because there are three inductive methods which conflict with each other, and yet which all meet the requirements of coherence. (But see Salmon's refutation of this claim, discussed above). The three methods, informally, are these: the more often you see  $p$ , the more often it will recur (standard method); the

<sup>13</sup> Ewing, "Causality and Induction," p. 466.

<sup>14</sup> *Ibid.*, p. 474.

<sup>15</sup> Freudenthal, "Abus philosophiques de la statistique."

<sup>16</sup> Burks, "The Presupposition Theory of Induction."

more often you find  $p$ , the less often it will recur (inverse method); and past and future occurrences of  $p$  are stochastically independent (random method). Burks argues that we can have no *inductive* ground for choosing among these three methods; to choose one is clearly to adopt a presupposition, an assumption.

In another article Burks goes on to show that the natural presupposition does not suffice for the justification of induction. There he adds two new presuppositions: (1) a presupposition concerning limited variety—that there are a finite number of irreducible, first-order, extensional, monadic, symbolic properties; and (2) a presupposition concerning uniformity—i.e., causality. This latter principle has two parts: (a) if one substitution instance of a causal universal is true, then the causal universal is true; and (b) there is a region in which some causal connections hold. It is not hard to see that the generality of these presuppositions precludes their use in the justification of any particular inductive inferences (as usual for presuppositions); and even for the general justification of induction, they are too weak. "If we could justify the claim that, relative to the presuppositions, each causal universal under consideration is likely to be in the group of true ones as any other causal universal, we would have the result we are looking for. It is clear that our three presuppositions do not suffice to justify this claim, or any similar claim that leads to the result we are looking for, and we must confess that we have not yet found a presuppositional statement which will do this."<sup>17</sup> There is no point, on Burks's view, in trying to pretend that these presuppositions, if and when they are properly formulated, are anything but presuppositions, that is to say, bare-faced assumptions. To say that they are probable only (as Keynes said of his Principle of Limitation of Natural Variety) is just as synthetic an assertion as to say that they are true.

John O. Wisdom presents four postulates for induction<sup>18</sup> that are much like those offered by Russell. They are: (1) the limitation of natural variety, (2) spatio-temporal proximity, (3) the uniform generation of properties, and (4) the

principle of small and great changes. Principles (1) and (3) are self-explanatory; (2) assures us that things don't change magically for no reason; and (4) asserts the possibility that small changes (e.g., the addition of a neutron to one nucleus) can have profound effects. Where Russell takes his postulates to be premisses that must be believed if science is to be sound, and Burks regards his presuppositions as premisses which (whether they are accepted or not) would suffice to render inductive reasoning sound, for Wisdom, "... the four inductive principles constitute no more than a description of the universe where scientific conclusions have in fact held good. . . ."<sup>19</sup>

J. G. Kemeny, too, professes to find factual assumptions underlying science. "In short, induction cannot be justified. We can only base it on a more or less plausible sounding assumption. . . . We may not be able to justify the assumption, but we must have faith in some such assumption if life is to be possible."<sup>20</sup> W. Donald Oliver also claims that "... we need to invoke certain general assumptions about the nature of the subject matter of science."<sup>21</sup> H. G. Alexander arrives at much the same conclusion.<sup>22</sup>

None of these recent writers attempts to justify his assumptions. Such a justification would be a justification of induction, and that, as Hume showed, is not to be had in the sense that these writers desire it. For none of them is the truth of their postulates more than a pious wish. As N. Goodman says, the "course of accepting an unsubstantiated and even dubious assumption much more sweeping than any actual predictions we make seems an odd and expensive way of justifying (inductions)."<sup>23</sup> (Wisdom could escape this condemnation, for he does not ask us to *accept* his principles—or does he, perhaps indirectly?) Furthermore, the assertion that these postulates will actually help us to reconstruct a specific piece of scientific argument on a sound basis is highly dubious. The arguments offered by the authors of such postulates are invariably of a highly abstract character, and the more plausible a postulate is made to seem, the farther into abstraction it appears to be removed from living science.

<sup>17</sup> Burks, "On the Presuppositions of Induction," p. 597.

<sup>18</sup> Wisdom, *Foundation of Inference in Natural Science*.

<sup>19</sup> *Ibid.*, p. 162.

<sup>20</sup> Kemeny, *A Philosopher Looks at Science*, p. 121.

<sup>21</sup> Oliver, "A Re-Examination of the Problem of Induction," p. 778.

<sup>22</sup> Alexander, "Convention, Falsification, and Induction."

<sup>23</sup> Goodman, *Fact, Fiction, and Forecast*, p. 65.

## III.

The principles and postulates that have been offered for the justification of induction have been criticized on two grounds: the difficulty of accepting them, and their impotence in providing grounds for actual scientific arguments. One of these two criticisms is met by another approach to inductive logic, that which takes the principle of induction itself (or something presumably equivalent to it) as analytic. (A special and elaborate case of this is the taking of a logical theory of probability as the foundation of inductive logic. I shall return to this approach later.)

It has been maintained that the uniformity of nature is a tautology, on the grounds that if things started acting very oddly, we would not decide that nature was no longer uniform, but simply that we were dreaming and no longer observing nature.<sup>24</sup> This gambit, like that of K. Campbell<sup>25</sup> who calls our attention to the fact that continuing concomitances are *presupposed* in our language and thought, simply won't do what is expected of it. Even if the existence of uniformities were enough to enable us to justify induction—which it is not, since we can only justify induction through uniformity when we have some further indication as to where the uniformities are and how far they extend—the analyticity of the uniformity principle would prevent us from using it to justify extrapolations into the future. If the uniformity of nature is made analytic, the continuing existence of nature cannot be regarded as analytic, but becomes contingent, and is precisely what inductive justification would have to depend on.

It is more plausible to maintain that the principle of induction itself is analytic, and at the same stroke to justify the belief that past uniformities will continue, and that nature, as we know it, will continue to exist. The first explicit statement that I know of the thesis that the principle of induction is analytic is by Asher Moore.<sup>26</sup> The principle of induction which he formulates is "It is more probable than not that uniformities, either universal or statistical, which have been observed to hold

uniformly in the past experience will continue to hold uniformly in the future."<sup>27</sup> This principle, he says, is analytic; it is analytic of what we mean by "probable" or "reasonable." (It goes without saying that the word "probable" here does not carry any commitment concerning future frequencies.) Against this view, May Brodbeck<sup>28</sup> argues that by making the principle of induction analytic, we find ourselves with no basis on which to construct an inductive logic. What is the basis for preferring large samples to small ones? Highly varied ones to homogeneous ones? According to her, all such knowledge is based on induction. Furthermore, the real problem of justification is untouched: to say that we have observed uniformity in the past and that this *analytically* implies that it is reasonable to expect uniformity in the future is to say no more than that we have observed uniformity in the past. "Yet if there is any problem of induction, it is about the relationship between the observed and unobserved . . ."<sup>29</sup> which is not touched at all by an analytical principle of induction. Moore answers that "The rules of induction, concerning mixing, large samples, varied data, and so on, 'constitute the full expansion of the Principle of Induction itself.'"<sup>30</sup> He neither develops an inductive logic nor discusses in any detail Brodbeck's most serious contention, that the problem of induction concerns the relationship between the observed and the unobserved, and that an analytic principle can have nothing to contribute to the elucidation of this relationship.

Grover Maxwell does attack that problem, maintaining that "It is necessarily true that the future will to some extent resemble the past."<sup>31</sup> His argument for this is that we can ask questions only within a conceptual framework; to doubt that crows will be black in the future as they have been in the past, presupposes that crows will remain an identifiable species, and that there will be this much uniformity in the future at least; " . . . unless it is presupposed that the future will to some extent resemble the past, we cannot meaningfully ask questions about the future at all."<sup>32</sup> It does not seem to me that this follows at all: we are,

<sup>24</sup> Brown, *Probability and Scientific Inference*, p. 19.

<sup>25</sup> Campbell, "One Form of Skepticism about Induction."

<sup>26</sup> Moore, "The Principle of Induction."

<sup>27</sup> *Ibid.*, p. 741.

<sup>28</sup> Brodbeck, "An Analytic Principle of Induction."

<sup>29</sup> *Ibid.*, p. 750.

<sup>30</sup> Moore, "The Principle of Induction (II): A Rejoinder to Miss Brodbeck," p. 752.

<sup>31</sup> Maxwell, "An 'Analytic' Vindication of Induction," p. 43.

<sup>32</sup> *Ibid.*, p. 44.

after all, asking the questions now; that the questions will not be meaningless in the future is still open to question; and what is even more serious from the point of view of this article is the fact that even an answer to that question would not lead us to any principles of inductive logic for the justification of particular inductions on the basis of particular evidence. The important question is, "In *what respects* will the future resemble the past?" Observe that in this respect Maxwell's thesis is only a slightly attenuated form of Campbell's.

H. A. Nielsen<sup>33</sup> develops a similar principle for statistical inference, arguing that arguments proceed, not from a *subclass* to a population, but from a *sample* to a population. The latter inference is sound, because *fairness* is a logical requirement of a sample. Thus, according to Nielsen, a sample *must* be like the population from which it was drawn. (Nielsen is wrong in supposing this—a certain proportion of fair samples will be *unlike* the population from which they are drawn.) The problem remains, of course, that it is difficult to tell when a particular subclass is a *sample* in Nielsen's technical sense.

Some writers have maintained the analyticity of the principle of induction in a less sweeping (and more plausible) form. Thus, N. R. Hanson<sup>34</sup> maintains that nothing but reflection is needed to discover the goodness of reasons. G. Buchdahl says, "There is a logic built into our language . . . such that we view testing procedures [*sic*] as evidence for universal laws, naturally described as inferring techniques."<sup>35</sup> (Hanson, too, wants to think of the results of induction as "inference tickets"—Ryle's term—rather than as universal propositions. I shall have comments to make on this thesis when I deal with the pragmatic approach to induction.)

M. Black and Goodman adopt a similar point of view with respect to induction. Black, although he states flatly that he rejects "the view that induction needs any 'justification' or 'vindication' in the senses which philosophers have usually attached to these words,"<sup>36</sup> is perfectly willing to talk of the

"unsolved problem of a satisfactory formulation of canons of inductive inference. . . ."<sup>37</sup>

Goodman writes, "An argument that [conforms to the general rules of deductive inference] is justified or valid, even if its conclusion happens to be false. . . . Analogously, the basic task in justifying an inductive inference is to show that it conforms to the general rules of induction."<sup>38</sup> Goodman shows that the solution of the problem of spelling out these rules depends on the prior solution of the problem of projectibility, to which I shall come back in due course; this is a problem that plagues nearly all the formalistic approaches to the construction of an inductive logic.

Another writer in this vein is Jerrold Katz,<sup>39</sup> who claims that the traditional search for an *a priori* inductive logic is misconceived, and should be replaced by a factual examination of the reasons for which scientists do accept those hypotheses they accept.

#### IV.

One of the defects of all the "ordinary language" approaches is that they do not make the logical structure of these "good inductive reasons" explicit. N. Rescher has worked on this problem; his results appear in two recent papers. In a brief paper in *Analysis*<sup>40</sup> he tries to formalize the concept of plausible implication. This formalization he regards as ". . . a step towards the formulation of a logical theory of inductive modes of reasoning." In an earlier paper,<sup>41</sup> he attempted to formulate a theory of evidential support. In both of these papers, Rescher presupposes the availability of a logical measure function  $L$  defined for sentences  $p$  and  $q$  in a definite formalized language. He defines the *degree of evidential support* that  $q$  gives to  $p$ ,  $des(p, q)$  as:

$$\frac{L(p, q) - L(p)}{1 - L(p)} L(q).$$

A statement  $q$  is confirming evidence for  $p$ , if (1)  $q$  renders  $p$  more likely than not, and also (2)  $q$  renders  $p$  more likely than before. Like the authors mentioned above, he feels that the inductive

<sup>33</sup> Nielsen, "Sampling and the Problem of Induction."

<sup>34</sup> Hanson, "Good Inductive Reasons."

<sup>35</sup> Buchdahl, "Convention, Falsification, and Induction."

<sup>36</sup> Black, *Models and Metaphors*, p. 262.

<sup>37</sup> *Ibid.*, p. 211.

<sup>38</sup> Goodman, *Fact, Fiction, and Forecast*, p. 66.

<sup>39</sup> Katz, *The Problem of Induction and Its Solution*.

<sup>40</sup> Rescher, "Plausible Implication."

<sup>41</sup> Rescher, "Theory of Evidence."

problem is to "analyze and codify the rules of evidence in various domains. . . ." <sup>42</sup> He uses his definitions to lay down some general rules of evidence, applicable to all domains. For example, he shows that a statement that entails another is evidence for it; and that a component of a conjunction is evidence for it. He also points out "illicit rules"—for example, that it is not the case that if  $p$  is evidence for  $q$ , and  $r$  entails  $p$ , then  $r$  is evidence for  $q$ .

Another proposal along the same lines has been made by Kemeny and Oppenheim. <sup>43</sup> Again a logical measure function  $m$ , giving each state description in a definite language a weight, is presumed to be on hand. The technique is analytic, where Rescher's technique was synthetic. Kemeny and Oppenheim lay down ten conditions which an adequate theory of factual support should, on their view, satisfy. The fourth condition, for example, concerns equivalent hypotheses and equivalent evidence: If  $\vdash H \equiv H'$  and  $\vdash E \equiv E'$ , then  $F(H, E) = F(H', E')$ , i.e., the degree of factual support given  $H$  by  $E$  will be the same as that given  $H'$  by  $E'$ . The sixth condition stipulates that  $F$  is to be a function of the following measures only:  $m(H, E)$ ,  $m(H, \bar{E})$ ,  $m(\bar{H}, E)$  and  $m(\bar{H}, \bar{E})$ . Having laid down the conditions of adequacy, they look for the simplest function of the measures mentioned in the sixth condition that will satisfy all the other conditions as well. The result is surprising and interesting. The function  $p$  which assigns to each state description the same probability (the "Wittgensteinian" measure function) is introduced as an auxiliary function; it is related to the general function  $m$  by a certain number  $q$ . In the final formula, if the  $m$ -functions are expressed in terms of  $q$  and the  $p$ -functions, all the  $q$ 's cancel out, and we are left with:

$$F(H, E) = \frac{p(E, H) - p(E, \bar{H})}{p(E, H) + p(E, \bar{H})}$$

for the degree of factual support lent  $H$  by the evidence  $E$ .

Popper, whose general theory of scientific logic we shall consider later, also offers a number of formulas, expressed in terms of logical probabilities, for degree of corroboration. They are presented in a series of three notes in the *British Journal for the Philosophy of Science*. <sup>44</sup> In the first of these articles he defines the degree of explanatory power of  $x$ , with respect to  $y$ ,  $E(x, y)$  as:

$$E(x, y) = \frac{P(y, x) - P(y)}{P(y, x) + P(y)}$$

This definition is very nearly the same as that used by Kemeny and Oppenheim for degree of factual support. For Popper it is only a starting point; he then defines the degree of corroboration of  $x$ , given by  $y$ ,  $C(x, y)$ , as:

$$C(x, y) = E(x, y) (1 + P(x) P(x, y))^{45}$$

and shows that this function satisfies all of his conditions of adequacy for degree of corroboration.

Both the concept of explanatory power and the concept of degree of corroboration may be relativized to a statement (of background knowledge, for example)  $z$ .

$E(x, y, z)$ , the explanatory power of  $x$  with respect to  $y$  in the presence of  $z$ , is defined as:

$$\frac{P(y, xz) - P(y, z)}{P(y, xz) + P(y, z)}$$

and  $C(x, y, z)$ , the degree of corroboration of  $x$ , with respect to  $y$ , in the presence of  $z$ , is defined as:

$$C(x, y, z) = E(x, y, z) (1 + P(x, z) P(x, yz))^{46}$$

Popper later discovered that his desiderata could be satisfied by a simpler formula (first stated in the *British Journal for the Philosophy of Science*, vol. 5 [1954-55], p. 334): <sup>47</sup>

$$C(x, y) = \frac{P(y, x) - P(y)}{P(y, x) - P(xy) + P(y)}$$

<sup>42</sup> *Ibid.*, p. 94.

<sup>43</sup> Kemeny and P. Oppenheim, "Degree of Factual Support."

<sup>44</sup> Popper, "Degree of Confirmation," "A Second Note on Degree of Confirmation," and "A Third Note on Degree of Corroboration or Confirmation." These three papers are reprinted as Appendix IX\* in Popper, *Logic of Scientific Discovery*, and page references will be made to that.

<sup>45</sup> *Ibid.*, p. 400.

<sup>46</sup> *Ibid.*

<sup>47</sup> *Ibid.*, p. 401.

<sup>48</sup> *Ibid.*

<sup>49</sup> Not, as stated on p. 400 of the *Logic of Scientific Discovery*, p. 359; observe also that the formula is given there incorrectly, with  $P(y, z)$  being written for  $P(y, x)$ .

In the second article, Popper considers the possibility of using a logarithmic measure of content, as had been suggested by both Kemeny<sup>50</sup> and C. L. Hamblin.<sup>51</sup> The third article is concerned primarily with the use in statistical inference of the concepts defined in the first article.

I. Levi argues<sup>52</sup> that considerations of epistemic utility (in particular, the principle that the utility of accepting a statement when it is true be independent of the evidence for the statement) are violated by Popper's notion of degree of corroboration, if that is taken as a guide to acceptance of scientific hypotheses. Instead, Levi offers the very simple utility assignments  $P(\sim x)$  and  $-P(x)$ , where  $P$  is the logical probability measure. Our epistemic expectation in accepting  $x$  on the evidence  $y$  is thus

$$U(x, y) = P(x, y) P(\sim x) - P(\sim x, y) P(x)$$

H. A. Finch has also dealt with the problem of the relation between the confirming power of observations among hypotheses.<sup>53</sup> He works from the point of view of modern decision theory (J.

Neyman, R. A. Fisher) and from the point of view of information theory and comes to the conclusion that the crucial quantity in confirming power is  $\frac{P_r(o/h) - P_r(o)}{P_r(o)}$ ; this is the measure of the power of  $o$  to confirm or disconfirm  $h$ .

Just to make the picture complete, I should also mention Carnap's measure of *increase of confirmation* of  $h$ , by  $e$ , as against the *a priori* probability of  $h$ ; he regards  $c(h, e) - c(h)$  as the suitable measure.

The following table will show the extent to which the intuitions of these writers coincide. All of the proposed measures turn out to be directly proportional to either the difference between the probability of the evidence  $e$  on the hypothesis  $h$  and the prior probability of  $e$ , or the difference between the probability of the evidence  $e$  on the evidence  $h$  and the probability of  $e$  on the denial of  $h$ ; and almost all the measures are inversely proportional to the prior probability of  $e$ .

Far and away the most thoroughgoing attempt to set out formal canons of inductive inference is that made by Day in his recent book.<sup>54</sup> Although

TABLE

Author	Proposed Measure of Corroboration or Evidential Power	Notation Rendered Uniform Simple Substitutions Made
Carnap	$c(h, e) - c(h)$	$P(h) \left[ \frac{P(e, h) - P(e)}{P(e)} \right]$
Rescher	$des(p, q) = \frac{L(p, q) - L(p)}{1 - L(q)} L(q)$	$\frac{P(h)}{P(\bar{h})} [P(e, h) - P(e)]$
Kemeny-Oppenheim*	$F(H, E) = \frac{P(E, H) - P(E, \bar{H})}{P(E, H) + P(E, \bar{H})}$	$\frac{[P(e, h) - P(e, \bar{h})]}{P(e, h) + P(e, \bar{h})}$
Popper 1	$C(x, y) = \frac{P(y, x) - P(y)}{P(y, x) + P(y)} (1 + P(x)P(x, y))$	$1 + P(h)(P(h, e) - P(e)) \left[ \frac{P(e, h) - P(e)}{P(e, h) + P(e)} \right]$
Popper 2	$C(x, y) = \frac{P(y, x) - P(y)}{P(y, x) - P(x, y) + P(y)}$	$\frac{P(e, h) - P(e)}{P(\bar{h}) P(e, h) + P(e)}$
Levi	$U(x, y) = P(x, y) P(\sim x) - P(\sim x, y) P(x)$	$P(h) P(\bar{h}) \frac{P(e, h) - P(e, \bar{h})}{P(e)}$
Finch	$\frac{Pr(o/h)}{Pr(o)} - 1$	$\frac{P(e, h) - P(e)}{P(e)}$

\* (Notice here that  $P(e) = P(h)P(e, h) + P(\bar{h})P(e, \bar{h})$ , so that the denominator is just the sum of the unweighted probabilities of  $e$  on  $h$  and  $\bar{h}$ .)

<sup>50</sup> Kemeny, "A Logical Measure Function," p. 297.

<sup>51</sup> Hamblin, C. L., unpublished thesis referred to by Popper.

<sup>52</sup> Levi, "Corroboration and Rules of Acceptance."

<sup>53</sup> Finch, "Confirming Power of Observations Matricized for Decisions Among Hypotheses."

<sup>54</sup> Day, *Inductive Probability*.

the book has been called a "morass of verbiage,"<sup>55</sup> it represents a serious and sincere attempt to come to grips with the complexity of inductive argument. The basic point of view from which Day starts is Kneale's—the fundamental relation of inductive logic is the probabilification relation. But where Kneale denies the possibility of formalization, Day sets out to provide it. He still denies—as Kneale does—that a *system* of inductive logic is possible.<sup>56</sup> I am not sure what the function of formalization without systemization is; but apparently probabilification judgments are not related in an over-all system.

Day considers three kinds of probability; causal which is always based on statistical knowledge; inductive, in which the probabilifying proposition is asserted; and hypothetical, in which the probabilifying proposition is merely contemplated. Inductive Probability, for Day, is an evaluative concept, and while it is not a metrical concept, it admits of degrees. In fact, it seems to admit of precisely three degrees: "Thus  $\gamma P(p)$ ,  $\beta P(p)$  and  $\delta P(p)$  are read respectively 'It is very probable that  $p$ ,' 'It is probable that  $p$ ' and 'It is fairly probable that  $p$ .'"<sup>57</sup> No other Greek letters are introduced, and it does not seem that these three are intended as variables. On the other hand, he says that the degree of probabilification "varies directly as" some things<sup>58</sup> and "inversely as" others,<sup>59</sup> which would suggest a continuum of degrees of the probabilifying relation. The main part of the book is taken up with the discussion of particular probabilification formulas. For example, formula (10') reflects the structure of a "non-numerical proportional subject-predicate"<sup>60</sup> type of judgment:

$$(10') \quad \gamma PF(fMg, f'Mg, \gamma C_2, \gamma C_6).$$

Translation: the fact that most observed  $f$  are  $g$  ( $f'Mg$ ) and are varied ( $\gamma C_2$ ) and that this proportion is stable ( $\gamma C_6$ ) makes it probable ( $\gamma$ ) that most  $f$  are  $g$  ( $fMg$ ). Rule (2), concerning this formula reads, "A proportional subject-predicate generalization is more probabilified as the observed sample is more varied."<sup>61</sup>

Whether all this formal complexity serves any useful function is doubtful, but in any case there is a more serious objection to Day's philosophy of induction. It is that he supposes that he can con-

sider one definite piece of probabilifying evidence at a time, and then in some fashion compound probabilifying judgments to arrive at a general judgment based on a whole body of knowledge. At least I assume that the compounding must be possible, for it is clearly only relative to a whole body of knowledge, and not to one piece of evidence at a time, that we make inductive judgments. Lacking a systematization of probabilifying judgments, there is no way of doing this. W. Salmon<sup>62</sup> has pointed out an even more serious blunder. Throughout the whole work, Day has assumed that such relations as "probabilifies," "is relevant to," etc., are transitive; but it is well known that even for probabilities that are based directly on statistical information, this is not the case: it is perfectly possible that  $p$  makes  $r$  probable, that  $r$  makes  $q$  probable, and yet that  $p$  does not make  $q$  probable. Salmon also pointed out that a number of the formulas offered by Day lead to results so counter-intuitive that they had best be called simply false.

Although S. Toulmin<sup>63</sup> denies that significant arguments of any sort are analytic, he really has a theory of induction which places him in the same category as the writers we have been considering. Arguments may use either deductive or inductive warrants to get from their premisses to their conclusions; there is no real difference, for neither type of warrant yields a conclusive argument, and neither type of warrant (with the exception of the type of warrant used in mathematics) is strictly analytic. Some inductive arguments are warrant-establishing, as well as warrant using, while deductive arguments are not warrant-establishing. To find out what forms of argument are acceptable in physics, we must look at the ways physicists argue. "There is no explanation of the fact that one sort of argument works in physics, for instance, except a deeper argument also within physics. (Practical logic has no escape route . . . into the *a priori*.)"<sup>64</sup> No analysis of the strength of arguments is possible for Toulmin, because he regards "Probably" simply as a modal term, which indicates the speaker's *attitude* toward the proposition in question. To say that it will probably rain is simply to say that it will rain, but to say it guardedly.

<sup>55</sup> Hutten, review of Day.

<sup>57</sup> *Ibid.*, p. 40.

<sup>58</sup> *Ibid.*, p. 108, rule 5.

<sup>61</sup> *Ibid.*, p. 98.

<sup>62</sup> Toulmin, *The Uses of Argument*.

<sup>56</sup> Day, *Inductive Probability*, p. 304.

<sup>58</sup> *Ibid.*, p. 106, rule 3.

<sup>60</sup> *Ibid.*, p. 324.

<sup>62</sup> Salmon, review of Day.

<sup>64</sup> *Ibid.*, p. 258.

P. F. Strawson takes a position much like Toulmin's. He argues explicitly against the existence of a formal inductive logic: "In fact, we can never describe the strength of evidence more exactly than by the use of such words as 'slender,' 'good,' 'conclusive,' . . ."<sup>65</sup> It is possible to answer the question "Will induction continue to be successful?" in the affirmative, because we have good evidence for this; but that induction is *reasonable* is simply analytic. Like Toulmin, he feels that the way to find out how inferences *should* be made in a particular field is to look at the way in which inferences *are* made in that field. Strawson argues that there is no way of justifying induction in general, although there are ways of justifying particular inductions.<sup>66</sup> Further, he regards the problem of finding a general justification unreal. It is sufficient that inductive beliefs, as Hume pointed out, are natural. Strawson challenges his critics to *try* to choose a counter-inductive policy. "If it is said that there is a problem of induction, and that Hume posed it, it must be added that he solved it."<sup>67</sup>

J. C. C. Smart agrees: "There are countless general truths which we just find that we know; we never went through a process of arriving at them. . . . Science never began with the programme of establishing such truths inductively from observation of particular instances."<sup>68</sup>

R. Harré follows the trail blazed by Strawson and Toulmin. Like Strawson, he regards sensible questions about induction to be of the form, "How is this particular procedure justified?" and the request for a general justification of inductive standards to be misconceived. "There are no general standards of reasonableness everywhere appealed to."<sup>69</sup> Like Toulmin, he takes many of the statements of science to formulate principles of inference, which do not, according to him, require inductive justification. (This is surely a curious doctrine: it suggests that if we call something a principle of inference, we don't care whether or not it will lead us astray, while if we consider the corresponding statement, we take seriously the possibility that it might be in error.)

Hanson also writes in this vein: "That *F* obtains is a good reason for *G*' is necessarily true, if it is true at all. Nothing but reflection is needed to

discover the goodness of reasons."<sup>70</sup> Hanson, too, distinguishes between the problem (he admits that it is a problem) of justifying inference-permits and the problem of justifying general statements. "'All *F*'s are *G*'s" does not, by itself, disclose whether it is up for scrutiny *qua* inference-permit or *qua* the factual support for that permit."<sup>71</sup>

The temptation to disregard completely the normative element that can be discerned in the appraisal of inductive arguments has been succumbed to by a number of writers. Strawson, Toulmin, and Harré—and to some extent Hanson—have all said that to evaluate inductive arguments we must look at inductive arguments that are actually regarded as valid in certain fields; Strawson even refers to Hume's observation that induction is a natural instinct. But far more extreme views of this sort are held by D. G. C. Macnabb, who claims that since induction is founded on instincts that we share with the higher animals, the way to understand it is to look to biology, rather than to logic,<sup>72</sup> and E. Gross, who writes, ". . . science is the culture of a society of scientists." Since "All cultures have (are) a set of behavioral prescriptions for satisfying needs,"<sup>73</sup> what we must do is to apply sociological methods to the culture in question in order to uncover the rules of inductive logic.

Toulmin's claim that it is profitable to look at certain classes of general statements as rules of inference has been examined carefully by H. G. Alexander.<sup>74</sup> He concludes that nothing is added to the traditional account in terms of enthymemes by talking instead of "material rules of inference," or "inference tickets." Alexander suggests that both deductive and inductive arguments be reconstructed in the traditional way, including all the material premisses explicitly as premisses, and that a deductive rule be added for passing from "Most *A*'s have been *B*'s" to "Probably this *A* is a *B*." I shall deal later on with the difficulties of the statistical syllogism. Here I should like only to observe that Alexander is wrong when he claims that nothing is gained by construing certain general material propositions as rules of inference; what is gained is the possibility of maintaining that induction can be justified inductively without circularity. Further discussion of the logical aspects

<sup>65</sup> Strawson, *Introduction to Logical Theory*, p. 247.

<sup>66</sup> *Ibid.*, p. 21.

<sup>67</sup> Harré, "Dissolving the 'Problem' of Induction."

<sup>68</sup> *Ibid.*, p. 127.

<sup>69</sup> Gross, "Toward a Rationale for Science," p. 838.

<sup>70</sup> Strawson, "On Justifying Induction."

<sup>71</sup> Smart, "Excogitation and Induction," p. 198.

<sup>72</sup> Hanson, "Good Inductive Reasons," p. 123.

<sup>73</sup> Macnabb, "Hume on Induction."

<sup>74</sup> Alexander, "General Statements as Rules of Inference."

of the curious doctrine of inference tickets, originally proposed, I believe, by G. Ryle<sup>75</sup>, may be found in H. W. Castañeda<sup>76</sup>, and J. C. Cooley.<sup>77</sup>

### V.

Some writers, convinced of the dishonesty of simply postulating certain material principles which will justify induction (as Russell says, such postulation has the same advantages as theft), and yet profoundly dissatisfied with an analytic principle of induction which does not give what they want—namely, some assurance that induction will often, if not always, lead to success—have seriously tried to show that the principle of induction can be justified without circularity on an inductive basis. One such attempt, that of R. B. Braithwaite, has been commented on by both A. Shimony and me, and found wanting. Braithwaite does not avoid circularity, according to Shimony; and that he certainly provides no persuasive inductive argument for the inductive principle is indicated by my counter-examples. But Braithwaite's argument was brief and incidental to the main line of argument in his book; it would be unwise and unfair to regard it as the last word in the inductive justification of inductive arguments.

Max Black has been one of the staunchest supporters of the non-circularity of the inductive justification of induction.<sup>78</sup> The possibility of constructing an inductive argument supporting a principle of induction which avoids circularity depends on the distinction between a *premiss* of an inference, and a *rule* in accordance with which one infers. The most important argument Black examines is this one:<sup>79</sup>

In most instances the use of  $R_2$  in arguments with true premisses examined in a wide variety of conditions,  $R_2$  has been successful.

Hence (probably)

In the next instance to be encountered of the use of  $R_2$  in an argument with a true premiss,  $R_2$  will be successful.

where  $R_2$  is: "To argue from 'Most instances of  $A$ 's

examined in a wide variety of conditions have been  $B$ ' to (probably) 'The next  $A$  to be encountered will be  $B$ '."<sup>80</sup>

Black argues that this argument cannot possibly be circular, since its conclusion is not one of its premisses, or even hidden in one of its premisses. Indeed, the conclusion does not even follow *deductively* from the premisses: the premisses could perfectly well be true and the conclusion false. Though this would be bad for a deductive argument, it saves our inductive argument from circularity. Furthermore such an argument may perfectly well be legitimate,<sup>81</sup> if we have gone to great lengths to satisfy ourselves of the propriety of using the rule we are using; and indeed the argument may also be "correct without triviality,"<sup>82</sup> for there is nothing to prevent us from taking as the evidential premiss in such an argument, material that we did not refer to in the examination of the legitimacy of the rule of inference in accordance with which we make the inference.

Peter Achinstein<sup>83</sup> offers two objections to the cogency of Black's argument. The first is a counter-example. Consider the rule of inference: Argue from "No  $F$  is  $G$ " and "Some  $G$  is  $H$ ," to the conclusion, "All  $F$  is  $H$ ." Call this rule of inference  $D$ . Now look at the self-supporting nature of the following argument:

*premiss*: No argument using  $D$  as its rule of inference is an argument which contains a premiss beginning with the term "all"; some arguments containing premisses beginning with the term "all" are valid.

*conclusion*: All arguments using  $D$  as their rule of inference are valid.

Black would maintain, of course, that there is no parallel here. There is a similarity of structure, to be sure, but while in the inductive case we know that the rule of inference is good—that is, we have "subjected it to scrutiny" and so on, in this case we know that the rule of inference is *not* good.

Achinstein's other argument is more telling; it is that "... to claim that a non-deductive rule of inference is correct or valid is to imply, ... that it will probably be successful in the next instance of its use. . . ." <sup>84</sup> If we call the self-supporting induc-

<sup>75</sup> Ryle, "Predicting and Inferring," and "If, 'So,' and 'Because'."

<sup>76</sup> Castañeda, "Are Conditionals Principles of Inference?"

<sup>77</sup> Cooley, "Toulmin's Revolution in Logic."

<sup>78</sup> Black, "Inductive Support of Inductive Rules."

<sup>79</sup> *Ibid.*, p. 197.

<sup>80</sup> *Ibid.*, p. 196.

<sup>81</sup> *Ibid.*, p. 203.

<sup>82</sup> *Ibid.*, p. 204.

<sup>83</sup> Achinstein, "The Circularity of a Self-Supporting Inductive Argument."

<sup>84</sup> *Ibid.*, p. 139.

tive argument (*a*), "... the assumption that (*a*) is valid ... does involve the claim that (*a*)'s conclusion is probable."<sup>85</sup> In a later note, Black agrees with this portion of Achinstein's argument.<sup>86</sup> According to Achinstein this makes (*a*) circular, since (*a*) is designed to show its conclusion probable. Now Black maintains that the conclusion of (*a*) is that *R* will be successful in its next application, and this is quite different from the assumption that *R* will probably be successful. Achinstein isn't satisfied by this gambit, and neither am I. I find myself asking what possible ground there can be for accepting the conclusion of the argument. The only ground that I can imagine being given is that *R* is probably a generally successful rule. Although, as Black points out, that *R* will probably be successful does not entail that it will be successful on its next application, it still seems to me that *R*'s probable success provides the *ground* for accepting the conclusion of the argument (*a*). There is a form of circularity here: we cannot get to know the adequacy, the correctness, of the *rule of inference* in accordance with which we construct the argument, without "simultaneously or antecedently"<sup>87</sup> getting to know the *ground* of the conclusion. A more careful analysis of "ground" would be required before we could spell out the type and degree of circularity involved in the simple inductive justification of induction. For our purposes here it really suffices that the inductive argument for the inductive principle has not been presented in such a way that all (or even most) philosophers find it cogent, and are willing to found their inductive logic on its basis.

What alternative ways are there of establishing the foundations of an inductive logic? One way, not unrelated to the foregoing attempt to achieve self-support, is to seek to justify the canons of inductive logic pragmatically. It is possible to argue, as Gilbert Ryle does,<sup>88</sup> that there is no inductive inference (and hence no inductive logic) and that the matter is settled as simply as this: we simply *see* that induction works. We do not, on Ryle's view, establish the general statement "All *A* are *B*" by observing *A*'s and *B*'s; we adopt the rule of inference: from the fact that something is

an *A*, infer that it is a *B*; and then we see from the success of the inferences made in accordance with this rule of inference that the rule *works*. This is not satisfactory. As already stated, there is no advantage, and there is considerable danger, in the attempt made by Strawson and Ryle, as well as by others, to deal with *material* rules of inference. In the second place, as Achinstein has argued, the problem is not to show that the rule (if that is how you wish to reconstruct the inference) has worked in the *past*; the problem is to show that it will continue to work. "The scientist not only points to the past successes of his theory, but he infers to its future successes as well."<sup>89</sup> He charges that Ryle is playing on the ambiguity of "it works." Ryle denies<sup>90</sup> that he is playing on the ambiguity of "works"; a general law or theory, he says, is like a design or a recipe: it is tenseless. There is no inference from successful inference to the truth of the *basis* of that inference any more than there is an *inference* from a successful soufflé to the validity of the recipe in accordance with which it was made. We *learn* the merits and demerits of a design or recipe by testing it—but to learn is not to infer.

The dispute is, I think, settled by Harry V. Stopes-Roe,<sup>91</sup> who says "... the fact is that, when one shows the goodness of a recipe or a theory by means of successful instances, there is a gap between what is literally shown—namely the successes to date—and the goodness of the recipe or theory." Whatever be the nature of this gap—and many would be happy to withhold the term "inference" from the process that leads us across it—it is that gap which an inductive logic or an examination of scientific method is called upon to fill.

A somewhat quantitative variant of the ordinary-language probabilistic approach is provided by Roy Harrod,<sup>92</sup> who is perfectly willing to compute *lower* bounds for probabilities, but who wants to reject the addition theorem and thus also most of the calculus of probability. (We shall later encounter writers who object to the multiplication theorem.) Harrod takes the following as the paradigm case of inductive reasoning: consider a man crossing an expanse. What is the probability that he will not encounter a boundary in the next

<sup>85</sup> *Ibid.*, p. 140.

<sup>86</sup> Black, "Self-Support and Circularity: A Reply to Mr. Achinstein."

<sup>87</sup> Black, *Problems of Analysis*, p. 198.

<sup>88</sup> Ryle, "Predicting and Inferring."

<sup>89</sup> Achinstein, "From Success to Truth," p. 8.

<sup>90</sup> Ryle, "Comment on Mr. Achinstein's Paper."

<sup>91</sup> Stopes-Roe, "Recipes and Induction: Ryle versus Achinstein."

<sup>92</sup> Harrod, *Foundations of Inductive Logic*.

part of his journey  $1/x$  as long as the distance he has already covered? Harrod's answer is "at least  $x/1 + x$ ," on the grounds that affirmative answers must (if the question were asked  $x$  times) be correct at least  $x/1 + x$  of the time.

There are two objections to be raised here. The first, a sophisticated one, has been raised by J. Bronowski:<sup>93</sup> the argument only holds for nice regular Euclidean-like spaces; it wouldn't hold for a space with more fence per pasture. My own objection would be that the "next"  $n/x$  feet, after I have traveled  $n$  feet, is not a random member of the set of  $x + 1$  intervals with respect to revealing a discontinuity, because I already have knowledge that there is no discontinuity in the first  $x$  intervals, so that if there is a discontinuity at all, it *must* be in the  $(x + 1)$ 'th interval. The probability that there is no discontinuity in the  $(x + 1)$ 'th interval, given that there is none in the first  $x$  intervals, is just the probability, *a priori*, that there is no discontinuity in the first  $x + 1$  intervals. We have no way to calculate this, but equally we have no way to be sure that it is as small as Mr. Harrod would like it to be.

Finally, although Harrod points out that his theory is designed only to provide a toehold for inductive inference, and is not intended as an explication of any of the methods of scientific inference that are actually used, he fails to give any clear indication of what these methods are, and of how they are related to his simple Euclidean induction.

## VI.

In 1950 Herbert Feigl published an article, "De Principiis non Disputandum . . . ?"<sup>94</sup> which has become one of the most mentioned pieces on the inductive problem; in it he distinguishes between *justification* and *vindication*. To ask for a general justification of inductive procedures is to ask the impossible, he argues. We can justify particular inductions by reference to general principles of induction, but we cannot justify the principles in the same way. What we can do, however, is to ask

for a vindication of the adoption of the rule of induction. Such a vindication consists precisely in showing that *if* the goal of predicting the future can be achieved, the rule of induction is a way to achieve it. (It is not necessarily the only way, or even the best way.) This was the kind of "justification" provided by H. Reichenbach for his "straight rule" of induction.<sup>95</sup> It is the kind of justification sought by many other writers—most notably in recent years by W. Salmon.

D. Kading argues that no pragmatic vindication succeeds, since, given the aim of predicting the future correctly, ". . . we have no proof that inductive procedure would be the best method of fulfilling this aim."<sup>96</sup> This argument misses the point: Feigl never claimed that induction was the *best* method. But both Black<sup>97</sup> and E. Madden<sup>98</sup> observe that in any *finite* run, we *can't* be sure that if anything will work induction will. This is the problem of the "short run." Reichenbach felt that it could not be solved directly.

John Lenz also mentions the "short run problem" as a problem for the pragmatic vindication of induction.<sup>99</sup> He also points to the fact that any asymptotic rule (e.g., estimate the relative frequency of  $A$ 's among  $B$ 's to be  $m/n + f(n)$ , where  $f(n)$  approaches 0 as  $n$  increases) allows the same justification as the conventional "straight rule"; and finally he argues that we have "no assurance that any of the predictions that science actually makes are correct or even probably correct."

Starting in 1956, Salmon has launched a heroic one-man attack on the opponents of the pragmatic vindication. Black argued<sup>100</sup> that the pragmatic justification could apply as well to the counter-inductive policy (if  $m$  of  $n$   $A$ 's have been  $B$ , predict a future ratio of  $\frac{n-m}{n}$ ) as to the conventional one (recall that Burks offered the same argument as one of the grounds for accepting his presupposition theory). This policy is self-corrective (that is, it reflects changes in our background knowledge), it may lead to success, and so on. Salmon shows<sup>101</sup> that this is not the case. Indeed, predictions made in accordance with Black's rule

<sup>93</sup> Bronowski, "The Scandal of Philosophy."

<sup>94</sup> Black, *Philosophical Analysis*.

<sup>95</sup> Reichenbach, "On the Justification of Induction."

<sup>96</sup> Kading, "Concerning Mr. Feigl's 'Vindication' of Induction."

<sup>97</sup> Black, "Can Induction be Vindicated?"

<sup>98</sup> Madden, "The Riddle of Induction."

<sup>99</sup> Lenz, "Problems for the Practicalist's Justification of Induction."

<sup>100</sup> Black, "The Justification of Induction."

<sup>101</sup> Salmon, "Regular Rules of Induction."

are contradictory: if  $\frac{1}{8}$  of the  $A$ 's are  $B_3$ ,  $\frac{5}{8}$  of them  $B_1$ , and  $\frac{3}{8}$  of them  $B_2$ , then Black's rule would lead to the prediction that  $\frac{7}{8}$  of the  $A$ 's are  $B_3$ ,  $\frac{3}{8}$   $B_2$ , and  $\frac{3}{8}$   $B_1$ , or (supposing that  $B_1$ ,  $B_2$ , and  $B_3$  are mutually exclusive) that  $\frac{1}{8}$  of the  $A$ 's have one or another of these properties. Salmon formulates the following two conditions which preclude such nonsensical results; the rules still allowed by them he calls *regular* rules, since their formal properties are those possessed by Carnap's regular  $c$ -functions.

Condition I: Given (1) a sequence  $S_i$  of events defined by the attribute  $A$ , (2) a set of properties  $B_1 \dots B_k$  which are mutually exclusive and exhaustive within  $S_i$  and (3) a sample of  $n$  members of  $S_i$ ,  $m_j$  of which have the property  $B_j$  ( $1 \leq j \leq k$ ); then, any rule  $R$  for making estimates  $E_j$  of the probability  $P(A, B_j)$  or of the relative frequency  $F(A, B_j)$  of  $B_j$  within the total sequence  $S_i$  must be such that

$$\sum_{j=1}^k E_j = 1.$$

Condition II: Under the circumstances detailed in Condition I,  $E_j$  must never be negative.<sup>102</sup>

Such irregular rules as Black and Burks have considered are thus shown to lead to contradiction.

Salmon attempted to deal with the short run problem about the same time<sup>103</sup> by suggesting that we make an estimate of the relative frequency, and then apply that estimate to the short run. As he realized soon afterwards<sup>104</sup> there are completely arbitrary rules—just as there are for inductive estimation generally—that admit the same justification that he offers for the natural rule.

Salmon provided a further development along these lines in 1961.<sup>105</sup> There he suggested two criteria—the criterion of convergence, and the criterion of linguistic invariance—which were offered in the hope that they would lead to plausible solutions both for the "short run" problem, and for the problem of selecting a *unique* member of the family of asymptotic rules. The criterion of linguistic invariance states that "no inductive rule is acceptable if the results it yields are functions of the arbitrary features of the

choice of language."<sup>106</sup> On this ground he rejects, for example, Carnap's theory of induction, since the degrees of confirmation reflect the richness of the language for which they are defined. In fact Salmon shows that every methodological rule of estimation with the possible exception of the straight rule will ultimately violate these two criteria. This is a large step toward a vindication of induction; it is much more than one would have believed it possible to prove. In the same paper he offers a solution to the short run problem—again, the natural one of making the short run estimate as close as possible to the estimate of the limiting frequency. Only this short run rule, of all possible asymptotic short run rules, appears not to violate the criterion of linguistic invariance.

Stephen Barker's comments<sup>107</sup> put this conclusion in a rather depressing light; for Barker shows, with no trouble at all, that the straight rule also violates the criterion of linguistic invariance. Consider Nelson Goodman's curious predicates "Grue" and "Bleen" with "Grue" meaning green and occurring before 2000 A.D. or blue and later than that; "Bleen" defined correspondingly. Clearly the straight rule, as applied to a sequence of emeralds, will lead us to make the estimate that the proportion of emeralds that are green is 1, and also that the proportion of emeralds that are grue is 1; but nothing can be both grue and green, so that we are led again to an inconsistency.

Salmon attempts to answer Barker's problem by stipulating that the straight rule of estimation be applied only to purely ostensive predicates.<sup>108</sup> He argues that Goodman type predicates are not "purely ostensive," because they must be defined—grue things don't *look* alike, while green ones do. A purely ostensive predicate is one which has these characteristics: "(1) it *can* be defined ostensively, (2) its positive and negative instances for ostensive definition *can* be indicated non-verbally, (3) the respect in which the positive instances resemble each other and differ from the negative instances is open to direct inspection."<sup>109</sup>

This solution is not altogether satisfactory. As the discussion following Salmon's paper showed, not all philosophers find the criterion of linguistic

<sup>102</sup> *Ibid.*, pp. 386–387.

<sup>103</sup> Salmon, "The Short Run."

<sup>104</sup> Salmon, "The Predictive Inference."

<sup>105</sup> Salmon, "Vindication of Induction."

<sup>106</sup> *Ibid.*, p. 246.

<sup>107</sup> Barker, "Comments on Salmon's 'Vindication of Induction'."

<sup>108</sup> Salmon, "On Vindicating Induction."

<sup>109</sup> *Ibid.*, p. 38.

invariance as natural as Salmon does.<sup>110</sup> Furthermore, even if we accept Salmon's solution, it involves abandoning a large part of what one took to be the program of inductive vindication. There remains, e.g., the problem of the short run. Black's main argument against Salmon is that the whole approach is wrong, because "Anything we observe in the short run is compatible with any value of the limiting frequency."<sup>111</sup>

Finally, this new result makes the second part of Barker's criticism of the earlier paper even more telling: "It remains to show how other types of inductive rules can be vindicated. It seems to me especially important to vindicate a rule for inferring from the limit of a relative frequency to short run relative frequencies and to introduce more complex rules of inductive inference to deal with the relation between scientific hypotheses and their evidence."<sup>112</sup> Here, for the moment, this approach to inductive logic rests.

## VII.

Salmon and Barker are among the few inductive logicians who take the "short run" problem seriously and see that short run predictions in the case of complete statistical knowledge are just as problematic as the inductive inference itself. The predictive inference forms the cornerstone of my own system of inductive logic.<sup>113</sup> I believe that if the short run problem can be solved, and if we can accept a rule of detachment for inductive conclusions, then most of our problems are over.

Salmon seems to have no qualms at all about applying a limiting frequency to a single event: "... a probability is assigned to a single event by referring it to a sequence of events, the selection of the sequence being governed by rather definite rules."<sup>114</sup> He does not state the rules. That rules are necessary is clear, as Barker has shown,<sup>115</sup> since the "proportional syllogism" so effectively em-

ployed by D. C. Williams<sup>116</sup> is not deductively valid, and its cogency can be destroyed by adding a premiss. G. Chatalian makes similar observations.<sup>117</sup> R. A. Fisher does make an attempt to state a rule: "The subject of a statement of probability must not only belong to a measureable set of which a known fraction fulfils a certain condition, but every subset to which it belongs, and which is characterized by a different fraction, must be unrecognizable."<sup>118</sup> I have shown that this rule is not sufficient and that any rule will have to be complex.<sup>119</sup> Henry Finch<sup>120</sup> has worked on the theory of proportionally quantified syllogisms, but he has dealt so far only with such proportional modifiers as "a few," "most," etc., and he does not deal with the problem of background knowledge. A. J. Ayer has told us to choose "the narrowest class in which the property occurs with an extrapolable frequency,"<sup>121</sup> but this merely makes the word "extrapolable" bear the whole unanalyzed problem of induction.

In my own system of inductive logic, the notion of randomness is fundamental (it is not taken as a primitive notion, however, but defined, and the statistical syllogism is validated by a randomness requirement). I define a sequence of rational corpora of different *levels* each level corresponding to a different notion of "practical certainty." The highest level of rational corpus corresponds to philosophical certainty and contains only logical and mathematical truths; and (for the purposes of constructing an inductive logic) certainties of sense. The language in which the statements are phrased as that of *ML*, supplemented by a finite number of primitive predicates, and axioms governing those primitive predicates. The probability of a statement of the form  $a \in A$  is the pair of numbers  $(p, q)$  when (roughly speaking)  $a$  is a random member of some set  $B$ , and the proportion of objects in  $B$  that belong to  $A$  is known to be in that interval  $(p, q)$ . By adopting the principle that if  $S$  is known to be equivalent to

<sup>110</sup> Kyburg, *Induction: Some Current Issues*.

<sup>111</sup> Black, "Comments on Salmon's Paper."

<sup>112</sup> Barker, "Comments on Salmon's 'Vindication of Induction,'" p. 54.

<sup>113</sup> Kyburg, *Probability and the Logic of Rational Belief*.

<sup>114</sup> Salmon, "The Short Run," p. 215.

<sup>115</sup> Barker, *Induction and Hypothesis*, p. 77.

<sup>116</sup> Williams, *The Ground of Induction* and "On the Direct Probability of Inductions."

<sup>117</sup> Chatalian, "Probability: Inductive versus Deductive."

<sup>118</sup> Fisher, *Statistical Methods and Scientific Inference*, p. 57.

<sup>119</sup> Kyburg, "Probability and Randomness."

<sup>120</sup> Finch, "Validity Rules for Proportionally Quantified Syllogisms."

<sup>121</sup> Ayer, "On the Probability of Particular Events," p. 370.

$T$ , the probability of  $S$  should be the same as the probability of  $T$ , we can, through the resources of the powerful *ML*, insure that since every statement is equivalent to one of the form " $a \in A$ ", every statement has a probability. Induction begins in the highest level rational corpus, where we have such statistical statements as: the proportion of 1000-member samples of  $A$  that have a relative frequency of  $B$  that is almost the same as the relative frequency of  $B$  in *all* of  $A$  lies between the limits .999 and 1.0. Relative to this highest level rational corpus, then, we can get such probability statements as "The probability is (.999, 1.0) that the proportion of all  $A$ 's that are  $B$ 's lies between .74 and .76."

The construction of rational corpora of lower levels proceeds by means of a rule of detachment: a statement is an ingredient of a rational corpus of level  $r_i$  if and only if the probability of that statement, relative to the rational corpus of next higher level, is at least  $r_i$ . Thus we would be entitled, on the basis of the above probability statement, to include the statement "The proportion of  $A$ 's that are  $B$ 's lies between .74 and .76" in the rational corpus of level .999.

The entire construction is done purely formally within Quine's protosyntax. Demonstrative induction—arguments from natural kinds, and the like—can be formalized perfectly adequately within the system. Furthermore the system provides a purely mechanical standard of cogency for inductive logic—not in the constructive sense mentioned by V. Somenzi<sup>122</sup>—but in the sense sought, for example, by Carnap.

One of the most crucial problems for any system that contains a rule of detachment for inductive conclusions is that of avoiding inconsistency. It is easy enough to think of a large set of statements each of which is *practically certain* in as strict a sense as you wish, but which is also such that their conjunction is certainly (or almost certainly) false. Consider, for example, the set of statements, "Ticket number  $i$  will not win the Irish Sweepstakes," and the conjunction of all such statements, which is equivalent to "No ticket will win the Irish Sweepstakes." It is clear that we must abandon complete logical closure for rational corpora—that  $S_1$  and  $S_2$  are so practically certain as to belong to a rational corpus should not mean

that their conjunction will be that certain. My system appeared to satisfy this condition; if  $S_1$  and  $S_2$  belong to a given rational corpus, there is no way of inferring that their conjunction does. I showed that their conjunction *did* belong to the rational corpus of next lower level, but this seemed unexceptionable.

But Fred Schick<sup>123</sup> showed that this held not only for two statements  $S_1$  and  $S_2$ , but for *any number* of them. This leads right back to the problem of the lottery, and, in view of the fact that I was able to show<sup>124</sup> that rational corpora were either consistent or empty, leads also to the conclusion that all the rational corpora of my system (except possibly the highest two) may be empty. Perhaps the difficulty could be dealt with by fixing up my definition of randomness, or by considering only *pairs* of rational corpora.

There is another problem that arises in my system as it did in Salmon's, and which appears to come up in every attempt to formalize induction: the problem of artificial Goodman-type predicates, such as "Grue" and "Bleen." The problem cannot be solved by considering only ostensive predicates, because it arises in the need to distinguish between arguments based on such logically true statistical premisses as: "Practically all 1000-member subclasses of  $A$  exhibit a proportion of  $B$ 's which differs from the proportion in  $A$  as a whole by less than  $\epsilon$ ," and "Practically all 1000-member subclasses of  $A$  exhibit a proportion of  $B$ 's which lies between  $r - \delta_1$ , and  $r - \delta_2$ , or between  $r + \delta_3$  and  $r + \delta_4$ , where  $r$  is the proportion of  $B$ 's in  $A$ ." In my book I solved the problem by the high-handed and *ad hoc* procedure of stipulating that only a certain set of expressions, constructed by certain techniques from the primitive predicates, were to be considered in the probability system. As between the two statistical statements just mentioned, for example, I chose to follow Neyman in basing my inference on a shortest confidence interval, as is obviously sensible. The logical basis of this preference, however, remains arbitrary.

So far, the peculiar predicates have resisted any general characterization. Goodman offered<sup>125</sup> a relative solution in terms of "entrenchment"—the degree to which a predicate is entrenched in our language. But might it not be that we could be speaking a different language? S. Barker and P.

<sup>122</sup> Somenzi, "Can Induction be Mechanized?"

<sup>123</sup> Schick, "Rationality and Consistency."

<sup>124</sup> Kyburg, "A Further Note on Rationality and Consistency."

<sup>125</sup> Goodman, *Fact, Fiction, and Forecast*.

Achinstein<sup>126</sup> attempt to show that we could not. Goodman rejects<sup>127</sup> their attempt to establish a logical asymmetry between "grue" and "green," and so does his student J. Ullian.<sup>128</sup> K. Small also tries to exhibit a significant asymmetry between "grue" and "green."<sup>129</sup> Salmon's solution (green things look alike, while grue things do not)<sup>130</sup> is like Goodman's (predicates that are well entrenched may be projected) in presupposing a principle of induction of the very sort they are designed to elucidate. Salmon's solution presupposes that things that look alike will continue to look alike, while Goodman admits that entrenchment in the past must be *assumed* a good guide to usefulness in the future.

Both H. Leblanc<sup>131</sup> and C. Hempel<sup>132</sup> have dealt recently with the problem created by Goodman's peculiar predicates. Leblanc carefully exhibits the inconsistencies into which one is led by the usual rules of inductive extrapolation, when one considers Goodman's peculiar predicates. He offers no solution to the problems that he raises. Hempel points out that the problem is not one of being able to determine which statements are lawlike and which are not (as Goodman claimed), for the same problem arises in connection with the extrapolation of numerical relations as quantitative laws: there are an infinite number of laws of the form  $y = f_1(x)$ ,  $y = f_2(x)$ , . . . which are mutually inconsistent, which are equally supported by the evidence, and which are equally lawlike. He too has no solution to offer, though both he and Leblanc point out that in formal theories of confirmation no *inconsistencies* are generated by the peculiar predicates. They nevertheless have unwelcome consequences.

It has been suggested—e.g., by Barker<sup>133</sup>—that this is a problem peculiar to the view that enumerative induction is basic to scientific inference. But it is clearly just as much a matter of concern to those who claim that scientific inference is a matter of selecting the simplest from a class of

acceptable hypotheses, for, if we spoke in a grue-bleen language, the hypothesis that all emeralds are grue would be much simpler (it would involve no reference to time, for example) than the hypothesis that they are green—i.e., grue until the year 2000, and bleen thereafter. The problem of finding some way of distinguishing between sensible predicates like "blue" and "green," and the outlandish ones suggested by Goodman, Barker, and others, is thus surely one of the most important problems to come out of recent discussions of inductive logic. It is also one of the most pressing, since only 36 years are left in which to solve it.

### VIII.

One of the most popular approaches to inductive logic is that followed, with variations, by Popper,<sup>134</sup> Barker,<sup>135</sup> Wisdom,<sup>136</sup> and others. It consists, first of all, in the denial that "induction" is fundamental to scientific inference. In its most extreme form, this denial is strengthened into a denial that there is any form of inference from experimental results to general scientific hypotheses, and that the contrary assertion is to be regarded as a regression into authoritarianism.<sup>137</sup> On this view of scientific method, what is fundamental is the free, unhindered construction of imaginative scientific hypotheses, on the one hand, and on the other, the subjection of these free creations to the most conscientious, rigorous, and severe tests we can devise. One of the most attractive features of this view is its simplicity. There is no need to worry about the relation between a scientific hypothesis and the data that suggested it, because the source of a hypothesis is utterly irrelevant. Whether I arrive at *H* by observing nature, by observing a crystal ball, or by a combination of deep breathing and free association, doesn't matter once the hypothesis is there. What matters is how well it stands up to tests. But the *logical* relationship between a hypothesis and the tests devised to test

<sup>126</sup> Barker and Achinstein, "On the New Riddle of Induction."

<sup>127</sup> Goodman, "Positionality and Pictures."

<sup>128</sup> Ullian, "More on 'Grue' and Grue."

<sup>129</sup> Small, "Professor Goodman's Puzzle."

<sup>130</sup> Salmon, "On Vindicating Induction."

<sup>131</sup> Leblanc, "That Positive Instances Are No Help," and "A Revised Version of Goodman's Paradox."

<sup>132</sup> Hempel, "Inductive Inconsistencies."

<sup>133</sup> Barker, "Rejoinder to Salmon."

<sup>134</sup> Popper, *The Logic of Scientific Discovery*.

<sup>135</sup> Barker, *Induction and Hypothesis*.

<sup>136</sup> Wisdom, *Foundations of Inference in Natural Science*.

<sup>137</sup> Agassi, "Corroboration Versus Induction."

it is straight-forwardly *deductive*: from *H* and boundary conditions and perhaps auxiliary hypotheses, a statement is deduced which can be directly tested. The relationship between a hypothesis and the tests we apply to it is *more* than logical, of course: the tests, to be significant, must be our sincere attempts to refute the hypothesis. This is a psychological matter, and offers nothing to the logician.

There are still problems for the logician to consider, however. For one thing, it is obvious that there will be, in any stage of knowledge, an infinite number of hypotheses that are not refuted by experience. It is true that we may not have thought of them all explicitly, but this seems to be accidental, and at any rate there is no doubt but that there are occasions when there are a large number of hypotheses we consider. (We can say that in considering functional generalization—when we seek the function that will relate pressure to temperature of a gas kept under constant volume, for example—there are an infinite number of possible laws that we are considering.) Two criteria have been suggested for making this choice: simplicity and content. According to some writers we should choose the simplest hypothesis (perhaps on the ground that it is most probable) while according to others we should choose the hypothesis that says the most (perhaps on the ground that it is most falsifiable, i.e., least probable!).

The most common view is that when we have to make a choice among unfalsified laws, we properly just choose the simplest. For the inductive logician, however, this merely raises the question as to what simplicity is and how it should be measured. A number of proposals have been made in recent years, ranging from the elegant proposal made by Svenonius<sup>138</sup> to the rather simple-minded sug-

gestion made by me.<sup>139</sup> Some of these proposals concern the logical simplicity of "primitive extra-logical vocabularies,"<sup>140</sup> and thus seem to have relatively little to do with the problems of inductive logic. The work of L. Svenonius, Goodman,<sup>141</sup> Kemeny,<sup>142</sup> and P. Suppes<sup>143</sup> falls into this category. My theory of simplicity takes it to be characteristic of a whole system of knowledge and takes as its basic measure the fewness of the number of quantifiers needed in the axioms and laws of the system. H. Jeffreys offers a definition of simplicity for functional laws (not for basic vocabularies) which depends on the number and type of parameters in the algebraic form of the law.<sup>144</sup> Barker proposes a measure of simplicity which uses Kemeny's logical measure functions<sup>145</sup> and which applies to theories: theory *S* is simpler than theory *T* if its logical measure is less than that of *T*, i.e., if for large enough *n*, *T* can be true of *n* things in more ways than *S* can.<sup>146</sup> Barker argues,<sup>147</sup> as does Jeffreys,<sup>148</sup> that we choose the simpler law not just because it is simpler, but also because it is "most probably true." This issue, and others relating to the use of the criterion of simplicity in induction, are discussed by R. Harrod,<sup>149</sup> P. G. Frank,<sup>150</sup> and Kemeny.<sup>151</sup> M. Scriven reminds us that simplicity is only one desideratum among others.<sup>152</sup> Harré introduces a distinction which might help to explain some of the conflicts about simplicity: the distinction between *formal* and *conceptual* simplicity.<sup>153</sup> Goodman and his followers have mainly been concerned with conceptual simplicity—the simplicity of the extra-logical basis of a language for science, while Jeffreys and most other writers have been concerned with the simplicity of various laws written within a *given* language. But as Harré also argues, the two kinds of simplicity are not logically independent.

The whole discussion of simplicity has been

<sup>138</sup> Svenonius, "Definability and Simplicity."

<sup>139</sup> Kyburg, "A Modest Proposal Concerning Simplicity."

<sup>140</sup> Goodman, "Recent Developments in the Theory of Simplicity," p. 429.

<sup>141</sup> Goodman, "New Notes on Simplicity."

<sup>142</sup> Kemeny, "Two measures of Complexity."

<sup>143</sup> Suppes, "Nelson Goodman on the Concept of Logical Simplicity."

<sup>144</sup> Jeffreys, *Scientific Inference*, p. 38.

<sup>145</sup> Kemeny, "A Logical Measure Function."

<sup>146</sup> Barker, *Induction and Hypothesis*, p. 180-182.

<sup>147</sup> *Ibid.*, p. 94, and Barker, "The Role of Simplicity in Explanation," p. 273.

<sup>148</sup> Jeffreys, *Scientific Inference*.

<sup>149</sup> Harrod, *The Foundations of Inductive Logic*.

<sup>150</sup> Frank, "The Variety of Reasons for the Acceptance of Scientific Theories."

<sup>151</sup> Kemeny, *A Philosopher Looks at Science*.

<sup>152</sup> Scriven, "The Principle of Inductive Simplicity."

<sup>153</sup> Harré, "Simplicity as a Criterion of Induction."

curiously inconclusive. Not only has there been no growing body of agreement concerning the measurement of simplicity, but there has been no agreement concerning the concept of simplicity for which we should seek a measure, or concerning the precise role that simplicity should play in the acceptance of scientific hypotheses.

Popper prefers to base the selection of a hypothesis from among those possible on a principle of falsifiability rather than on a principle of simplicity. We should accept that hypothesis which will be most quickly eliminated by tests if it is false, i.e., the most falsifiable hypothesis, the hypothesis with the greatest content. But he also shows<sup>154</sup> that falsifiability in this sense corresponds closely to our intuitive notion of simplicity. This correspondence leads to one of those curiously paradoxical sounding conflicts that seem more prone to rise in inductive logic than in any other field. The conflict is this: Barker and Jeffreys claim that, given a choice, we should select the most probable hypothesis, "most probable" corresponding with "simplest."<sup>155</sup> Popper, using a perfectly analogous concept of logical probability, claims the opposite, that we select the *least* probable (the most daring, the most falsifiable) hypothesis.<sup>156</sup> This conflict has been discussed by J. L. Harsanyi<sup>157</sup> who tends to side with Jeffreys in saying that we should prefer the hypothesis with the highest posterior probability. This misses the point for two reasons: first, Popper does not regard the notion of posterior probability (Carnap's "degree of confirmation") as a useful one; and second, the hypothesis with the highest posterior probability is simply that which describes what *has happened* (i.e., just describes the evidence) and leaves the whole future course of events undetermined. But this hypothesis would be selected by no one.

Indeed, the most interesting thing about this entire conflict is that neither Barker, Popper, Jeffreys, nor any other inductive logician would have any difficulty at all in deciding what hypothesis to accept in an ordinary experimental situation. This suggests that the very hypothesis "most probable" for Jeffreys is precisely that which is "least probable" for Popper. A part of the explanation is that suggested by Harsanyi: a law

which is most probable *a posteriori* may not have been most probable *a priori*. Other factors are these: Jeffreys is supposing that we have a list of possible *general* laws which might possibly govern occurrences of a certain sort. (Thus he does not even consider the laws Popper regards as "most probable"—i.e., laws going only slightly beyond our observations.) Furthermore, Jeffreys is using Bayes's theorem, according to which the probability of the hypothesis on the given evidence is equal to the probability of the evidence on the hypothesis (generally taken to be one), multiplied by the ratio of the *a priori* probability of the hypothesis to the *a priori* probability of the evidence. Disregarding the *a priori* probability of the hypothesis, we obtain a reconciliation of the two views by observing that to say that the hypothesis is highly falsifiable is to say that the *a priori* probability of the evidence is very small. But this is just what Jeffreys says: the posterior probability of the hypothesis is proportional to the reciprocal of the probability of the evidence. Even the ranking of laws in order of simplicity is roughly the same on the two views: a straight-line relationship is highly falsifiable (and highly improbable) on Popper's view and in general a theory involving a small number of parameters (by ruling out a large number of sets of possibilities) will be less probable than one involving a great number of parameters.<sup>158</sup> For Jeffreys, the theory with fewer adjustable parameters is both more probable and simpler. Other things being equal, then, the simplest law on both views is the one with the least number of parameters.

The major, and the only serious, conflict between the two views is that expressed in Popper's refusal to identify corroboration with posterior probability. This refusal is based on his denial that we choose the most probable hypothesis, and his assertion that we accept the hypothesis that has the greatest content, and which has stood up to severe tests and sincere attempts to refute it. Of course this conception of scientific acceptability raises certain problems, as Popper's critics have been quick to point out. D. C. Stove argues<sup>159</sup> that these psychological characterizations of what constitutes a test (it must be "sincere," an "attempt to falsify," etc.)

<sup>154</sup> Popper, *The Logic of Scientific Discovery*, Ch. vii.

<sup>155</sup> Jeffreys, *Scientific Inference*, p. 36.

<sup>156</sup> Popper, *The Logic of Scientific Discovery*, p. 141.

<sup>157</sup> Harsanyi, "Popper's Improbability Criterion for the Choice of Scientific Hypotheses."

<sup>158</sup> Popper, *The Logic of Scientific Discovery*, p. 380-381.

<sup>159</sup> Stove, review of Popper, *The Logic of Scientific Discovery*.

are quite irrelevant to the matter of evidential import: "To suppose otherwise would be to let differences of intention between two persons who subjected the same theory to the same test affect the question of the support, if any, which passing the test gives to the theory." Then there is the question of whether "test" is to be understood in a generic sense: if so, then we may ask, as G. J. Warnock does,<sup>160</sup> why passing one application of a test should convince us that a theory will pass subsequent applications of the same test; and if not, why we should believe that it will pass different tests. P. C. Gibbons<sup>161</sup> rejects the notion that there is any useful ordering of tests into more and less severe. Popper himself admits that "the requirement of sincerity cannot be formalized." It is possible to go one step further (for, as Buchdahl has remarked,<sup>162</sup> the notion of a "test" is ambiguous) and to assert that what it is for an observation to be a test of a theory cannot be formalized. This will become clear in the discussion of the paradoxes of confirmation; even J. W. N. Watkins admits that there are circumstances in which the observation of a white shoe constitutes a "test" of the theory that all crows are black. And R. H. Vincent<sup>163</sup> has shown that Popper's criteria, like every one else's, founder on Goodman's reef.

All this psychologism is distasteful to the logician who is looking for the timeless quality of evidential power. Furthermore, everyone recognizes differences in the degree to which hypotheses have been tested and stood up to tests. Not all advocates of the hypothetico-deductive method agree that the problem of distinguishing between well-supported hypotheses and highly speculative ones is a serious one, but Popper, for example, has taken it seriously enough to offer a definition of "degree of corroboration" to explicate the distinction.

### IX.

There is an approach to inductive logic which takes the distinction between well-confirmed and not-so-well-confirmed as fundamental. The central

proponent of this approach is Carnap, whose continuing researches into logical probability as the framework of an inductive logic have inspired considerable admiration and emulation in the past fifteen years. His original treatise of that time<sup>164</sup> exhibited certain shortcomings of which he himself was quite aware. Many of these shortcomings have been overcome by himself and his followers. The original system of logical probability was limited to first order functional calculi consisting of a finite number of individual signs and a finite number of one-place predicates that were assumed to be logically independent (this would exclude color predicates, for example, which could be assumed to be logically exclusive) and logically complete in the sense that anything to be said about the world could be said in terms of the predicates. These restrictions were quickly eliminated. Kemeny eliminated the requirement of logical independence<sup>165</sup> simply by stipulating (in the metalanguage) that the measure function  $m$  is to assign real numbers to all non-contradictory state descriptions so that the sum of the values is 1. At the same time, Y. Bar-Hillel pointed out that a problem is created for Carnap's system by the fact that relations generally have a logical structure<sup>166</sup>—a relation may be transitive, irreflexive, etc.—but this problem, though a real one for Carnap,<sup>167</sup> was soon dealt with by Kemeny for languages consisting of a first order functional calculus with properties, relations, and an identity sign. Later he published a more general system<sup>168</sup> applicable to functional calculi of all finite orders in which there was no restriction on the interdependence of the atomic sentences of the calculus, but only the following mild restrictions on the language itself:

1. The object language must be consistent.
2. It must not contain an axiom of infinity.
3. The number of types of individuals must be finite.
4. There must be a finite number of constants.
5. Each constant must be of finite order.

The measure function  $m'$ , reflecting the facts of logical interdependence, is obtained from a

<sup>160</sup> Warnock, review of Popper, *The Logic of Scientific Discovery*.

<sup>161</sup> Gibbons, "On the Severity of Tests."

<sup>162</sup> Buchdahl, "Convention, Falsification and Induction."

<sup>163</sup> Vincent, "Popper on Qualitative Confirmation and Disconfirmation."

<sup>164</sup> Carnap, *The Logical Foundations of Probability*.

<sup>165</sup> Kemeny, Review of Carnap, *Logical Foundations of Probability*.

<sup>166</sup> Bar-Hillel, "A Note on State-Descriptions."

<sup>167</sup> Carnap, "The Problem of Relations in Inductive Logic."

<sup>168</sup> Kemeny, "A Logical Measure Function."

measure function  $m$  defined for all state descriptions by the following relation:

$$m'(W) = \frac{m(A \cdot W)}{m(A)}$$

where  $W$  is an arbitrary statement and  $A$  is the conjunction of all the axioms reflecting logical structure in the language.

There have been a number of other workers in the same vineyard. R. M. Martin claimed to provide a simplification of the semantic basis of Carnap's theory.<sup>169</sup> Hilary Putnam has provided a definition for languages free of even Kemeny's restrictions.<sup>170</sup> Hugues Leblanc has provided a number of studies of various aspects of Carnapian logical probability.<sup>171</sup>

Several writers (Kemeny,<sup>172</sup> R. S. Lehman,<sup>173</sup> and Shimony<sup>174</sup> have found a justification for Carnap's criteria of adequacy in the consideration of bets as reflecting degrees of belief. Thus if  $P(A)$  is the probability of  $A$ , I should be willing to bet on  $A$  at the odds of  $P(A) : (1 - P(A))$ , and if  $P(\bar{A})$  is the probability of the negation of  $A$ , I should be willing to bet on  $\bar{A}$  at odds of  $P(\bar{A}) : (1 - P(\bar{A}))$ . But these are simply the opposite sides of the same bet, so we must have  $\frac{P(A)}{1 - P(A)} = \frac{1 - P(\bar{A})}{P(\bar{A})}$  or  $P(A) = 1 - P(\bar{A})$ . Relations sufficient to develop the whole of the probability calculus can be developed in this way. Recently Carnap has adopted this approach to the axioms of the calculus.<sup>175</sup> He has also been working with Kemeny and Jeffrey on the development of a theory of confirmation applicable to families of predicates (e.g., color-words).

One objection that has been raised against Carnap's theory (first by G. Bergmann), is the arbitrariness of the values of degree of confirmation. In the first place, the value may reflect the number of primitive predicates in the language; by adding a primitive predicate that does not

appear in either  $h$  or  $e$ , we may change the value of  $c(h, e)$ . And  $c(h, e)$  is also arbitrary in the sense that its value depends on the logical measure function with which we start, and the selection of one particular measure function from among the infinite number possible seems arbitrary. Kemeny suggested<sup>176</sup> making the arbitrariness explicit. Let  $e_0$  be the statement that  $s$  individuals all have each of  $t$  properties;  $h_0$  be the hypothesis that some other individual will have them too; and take  $c(h_0, e_0) = \frac{(s + k(\frac{1}{2})^t)}{s + k}$ . Taking  $k = 2^P$ , where  $P$  is the number of primitive predicates, yields Carnap's function  $c^*$ . This proposal is close to that made subsequently by Carnap himself.<sup>177</sup>

This arbitrariness is unsettling. Carnap suggests<sup>178</sup> that we can take the success of our inductive method in our actual universe as a guide to its suitability; but as J. W. Lenz has argued<sup>179</sup> that in choosing a particular  $c$ -function we must be either basing our decision on the hypothesis that  $c$ -functions that have been most adequate in the past will continue to be most adequate in the future (a synthetic hypothesis demanding justification), or we are basing our decision on some synthetic principle of induction. We do not need the principle of induction to justify probability statements; but we do need to make an assumption in choosing a  $c$ -function. "And where the principle needs to be assumed seems to me less important than that it has to be assumed."<sup>180</sup>

Carnap is still hoping to be able to find a confirmation function which will compel universal acceptance. "At the present time, I do not assert that there is only one rational  $C_{\sigma}$ -function."<sup>181</sup> Shimony has promised to look for an analytic principle of indifference which will provide the basis for an inductive logic.<sup>182</sup> Jeffreys, who also adopts a logical conception of probability, does not even regard the problem as serious; in the case of conflict among our intuitions, he says, we can

<sup>169</sup> Martin, "A Formalization of Inductive Logic."

<sup>170</sup> Putnam, "A Definition of Degree of Confirmation for Very Rich Languages."

<sup>171</sup> Leblanc, *Statistical and Inductive Probabilities*.

<sup>172</sup> Kemeny, "Fair Bets and Inductive Probabilities."

<sup>173</sup> Lehman, "On Confirmation and Rational Betting."

<sup>174</sup> Shimony, "Coherence and the Axioms of Confirmation."

<sup>175</sup> Carnap, "The Aims of Inductive Logic."

<sup>176</sup> Kemeny, "A Contribution to Inductive Logic."

<sup>177</sup> Carnap, "Remarks to Kemeny's Paper," and *The Continuum of Inductive Methods*.

<sup>178</sup> Carnap, *The Continuum of Inductive Methods*, pp. 75-77.

<sup>179</sup> Lenz, "Carnap on Defining 'Degrees of Confirmation'."

<sup>180</sup> *Ibid.*, p. 235.

<sup>181</sup> Carnap, "The Aims of Inductive Logic," p. 316.

<sup>182</sup> Shimony, "Coherence and the Axioms of Confirmation," p. 28.

always refer the problem of deciding on the *a priori* probabilities of laws to an international body of scientists.<sup>183</sup>

### X.

I shall return to the question of arbitrariness shortly: of more immediate concern are the arguments that have been presented purporting to show that any logical theory of probability, interpreted as explicating confirmation, is inherently paradoxical, or even contradictory.

To begin with, there is the classical paradox of confirmation. Consider the statement, "All crows are black." Anyone who believes in a logic of confirmation at all would regard the observation of a black crow as a (slight) confirmation of this statement. One of the most natural properties of confirmation is that if a sentence *S* confirms a sentence *T*, and *T'* is logically equivalent to *T*, then *S* confirms *T'* as well. Now "All crows are black" is logically equivalent to "All non-black things are non-crows," so the observation of a white shoe (which confirms the latter just as the observation of black crow confirms the former) will confirm "All crows are black." Finally the original hypothesis can be expressed: "Everything is either not a crow or else it is black"; and it is not hard to show that this hypothesis is confirmed by any object that is black (whether or not it is a crow) and any object that is not a crow (whether or not it is black). Thus as confirming instances for the law we have: a black crow, a white shoe, and a black cat.

Watkins recently offered this paradox as an argument against an inductivist view of the distinction between analytic and empirical statements.<sup>184</sup> In answer, Hempel admitted that these consequences of confirmation theory are "intuitively paradoxical," but insisted that they are "systematically unobjectionable." He also pointed out that perfectly analogous consequences arise on the Popper-Watkins theory of falsification. A white

shoe can perfectly well be regarded as the outcome of an attempt to falsify the theory that all crows are black: when I first looked I thought it was a crow, even though it was white, but then when I examined it, I saw that it was a shoe.<sup>185</sup> The argument was carried on by I. Scheffler<sup>186</sup> who attacked Watkins' arguments,<sup>187</sup> and finally pointed out that the disagreement may have stemmed from the fact that Hempel is not offering methodological prescriptions, while Popper is. Another exchange was instigated by D. C. Stove, who tried to throw some light on the argument by distinguishing between the pragmatic notion of a *test*, and the logical notion of *evidence*. The "attempt" to falsify the hypothesis that all crows are black by examining white shoes is obviously futile on anybody's theory, but "the attempts that might be made to falsify or instantially-confirm a hypothesis are quite irrelevant to the weight of the evidence, if any, resulting from such attempts."<sup>188</sup> A discussion ensued between Stove and the indefatigable Watkins.<sup>189</sup>

Alexander<sup>190</sup> also observed that the Watkins-Popper theory of falsification was subject to the same paradoxes as the theory of confirmation, but he pointed out that if we take account of our background knowledge we can save confirmation theory by looking at the matter quantitatively; following J. Hosiasson<sup>191</sup> he pointed out that since we know that most things in the world aren't ravens, the observation of a white shoe is not going to confirm "all ravens are black" as much as the observation of a black raven. J. Agassi jumped into the fray to save Popper and Watkins against Alexander, by observing (after a few polemical remarks about authoritarianism) that while a white shoe might be relevant on Popper's theory, as on Hempel's, it could be so only if it were observed as a result of a *test* of the generalization about ravens.<sup>192</sup> Watkins also contributed to the discussion of background knowledge,<sup>193</sup> and I. J. Good clarified some of the quantitative aspects of the "paradox."<sup>194</sup>

The discussion was admirably summed up by

<sup>183</sup> Jeffreys, *Scientific Inference*, p. 39.

<sup>184</sup> Watkins, "Between Analytic and Empirical."

<sup>185</sup> Hempel, "Empirical Statements and Falsifiability."

<sup>186</sup> Scheffler, "A Note on Confirmation," and "A Rejoinder on Confirmation."

<sup>187</sup> Watkins, *op. cit.* and "Professor Scheffler's Note."

<sup>188</sup> Stove, "Popperian Confirmation and the Paradox of the Ravens."

<sup>189</sup> Watkins, "Mr. Stove's Blunder"; Stove, "A Reply to Mr. Watkins"; Watkins, "A Reply to Mr. Stove's Reply."

<sup>190</sup> Alexander, "The Paradoxes of Confirmation."

<sup>191</sup> Hosiasson-Lindenbaum, "On Confirmation."

<sup>192</sup> Agassi, "Corroboration versus Induction."

<sup>193</sup> Watkins, "Confirmation Without Background Knowledge."

<sup>194</sup> Good, "The Paradox of Confirmation," and "The Paradox of Confirmation (II)."

J. L. Mackie<sup>195</sup> who traced the whole argument (overlooking only the Scheffler-Watkins exchange) and came to much the same conclusions that are apparent above. Given a complete lack of background knowledge, Hempel is perfectly correct, and a white shoe, a black cat, and a black raven will all confirm the generalization. Given some knowledge about the relative numbers of black objects and ravens and non-black objects and non-ravens, the bite may be taken out of the paradox by quantitative considerations, as Good, Hosiasson, and Alexander, show. And if we allow unlimited background knowledge to give meaning to the notion of a *test*, the observations of black ravens and non-black non-ravens confirm the generalization to a worth-while degree only if they are made in genuine *tests* of the generalization, and observations of black non-ravens never confirm it to a worth-while degree, in view of the fact that such an observation can never result from a genuine test.

There are three other paradoxes that are cited as arguments against confirmation theory. The first I shall consider is the paradox of ideal evidence, discussed both by Popper<sup>196</sup> and by R. H. Vincent.<sup>197</sup> Let  $a$  be the assertion that a particular toss of a given coin yields heads. It is clear that in the absence of any knowledge at all the *a priori* probability of  $a$  may plausibly be supposed to be  $\frac{1}{2}$ . Now let us subject the coin to extensive tests, and suppose that (say on the basis of a million tosses) we become very sure that the relative frequency of heads is  $\frac{1}{2} \pm \epsilon$ . Let the body of evidence for this assertion be  $e$ . Then the probability of  $a$ , given  $e$  is also  $\frac{1}{2}$ . Therefore the examination of a million tosses of the coin is utterly irrelevant to  $a$ . This is clearly paradoxical. The problem here is that of finding a way of taking account of the "weight of evidence," but according to Popper this cannot be done in view of the fact that "The fundamental postulate of the subjective theory [confirmation theory] is the postulate that degrees of the rationality of beliefs in the light of evidence exhibit a linear order."<sup>198</sup>

Although I myself regard this as a serious difficulty for most theories of degrees of confirmation, it is not as decisive as Popper makes it sound. These theories always contain, as an extra-formal

requirement, a principle of total evidence, which demands that if we have information about the behavior (or structure) of the coin, we use it; this may be taken as a principle demanding that we maximize the weight of evidence. It could also be maintained that there is no behavioristic way of distinguishing between the two situations Popper describes: I shall, in either case, be willing to wager on the truth of  $a$  at even money. I think, however, that a difference does arise when we consider  $A$ , the conjunction a set of five hundred statements like  $a$ . Given the evidence  $e$ , I will be very sure that very nearly half of the statements in  $A$  will turn out to be true. Without the evidence  $e$ , I will not be nearly so sure that nearly half the statements in  $A$  will turn out to be true. This difference in sureness will be reflected in my betting behavior.

The next paradox arises, like the classical one, from a conflict between our intuitive assessments of probability and our lightening calculations of related quantities. Vincent uses it to attempt to show that no plausible theory of confirmation can accept the multiplication axiom.<sup>199</sup> Let  $q$  consist of a hypothesis (such as Newton's laws) together with boundary conditions sufficient to entail  $p$ , where  $p$  is the statement, "A freely falling body near the earth will fall 144 feet in three seconds." Since  $q \rightarrow p$ , we have  $(q \cdot p) \leftarrow \rightarrow q$ , and (writing  $c(x, y)$  for the degree of confirmation of  $x$  given  $y$ ),  $c(p \cdot q, r) = c(q, r)$ . According to the multiplication axiom,

$$c(p \cdot q, r) = c(p, r) \cdot c(q, p \cdot r)$$

or

$$c(p, r) = \frac{c(q, r)}{c(q, p \cdot r)}$$

Since  $p$  cannot contribute much to the confirmation of  $q$ ,  $c(q, p \cdot r) \approx c(q, r)$  and  $c(p, r) \approx 1$ .

A numerical example shows that the conclusion is not valid, and that its paradoxical air results from thinking loosely about numbers. I borrow the example from Vincent, making it only slightly more specific. He says that if  $q$  is "All balls in urn  $U$  are red," and  $p$  is "the first three balls drawn are red," then by the above argument,  $p$  will be practically certain relative to our background knowledge. But assuming the equiprobability of structure descriptions and that we have five balls in the urn, we have:

<sup>195</sup> Mackie, "The Paradox of Confirmation."

<sup>196</sup> Popper, *The Logic of Scientific Discovery*, p. 407; "A Third Note on Degree of Corroboration or Confirmation."

<sup>197</sup> Vincent, "The Paradox of Ideal Evidence."

<sup>198</sup> Popper, "Third Note" reprinted in *Logic of Scientific Discovery*, p. 408.

<sup>199</sup> Vincent, "A Note on Some Quantitative Theories of Confirmation."

$$\begin{aligned} c(p, r) &= \frac{3}{10} \quad (\text{drawing with replacement}) \\ c(p, q \cdot r) &= 1 \quad c(q, r) = \frac{1}{5} \\ c(q, p \cdot r) &= \frac{3}{5} \end{aligned}$$

so that

$$c(p, r) = \frac{3}{10} = \frac{c(q, r)}{c(q, p \cdot r)} = \frac{\frac{1}{5}}{\frac{3}{5}}$$

To say that if the number of balls were larger, so that  $c(q, p \cdot r)$  were closer to  $c(q, r)$  we would find a paradox is false:  $c(q, r)$  and  $c(q, p \cdot r)$  will both decrease, but their ratio need not at all approach unity—indeed it may well *decrease*.

The last paradox is that regarded by Popper as a straight-forward contradiction in Carnap's system. The paradox is that:

"There are cases in which  $x$  is strongly supported by  $z$  and  $y$  is strongly undermined by  $z$ , while at the same time  $x$  is confirmed by  $z$  to a lesser degree than is  $y$ ."<sup>200</sup>

For example, if  $x$  is the statement that a six will turn up on the next throw of a die, and  $y$  is the statement that some other number than six will turn up, and  $z$  is the statement that an even number turned up, then  $z$  increases the probability of  $x$ , decreases that of  $y$ , while at the same time confirming  $x$  less than it confirms  $y$ :

$$\begin{aligned} c(x) &= \frac{1}{6}; c(x, z) = \frac{1}{3} \\ c(y) &= \frac{5}{6}; c(y, z) = \frac{4}{6} \end{aligned}$$

while

$$c(x, z) < c(y, z).$$

Popper claims that this is a self-contradictory state of affairs, for he claims that it is always self-contradictory to say that " $x$  has the property  $P$  . . . and  $y$  has not the property  $P$  and  $y$  has the property  $P$  in a higher degree than  $x$ ."<sup>201</sup>

This is a valid objection to Carnap's way of talking, for Carnap himself claimed that the statement " $a$  is warm and  $b$  is not warm and  $b$  is warmer than  $a$ " was self-contradictory, and the statement about confirmation sounds the same. In fact, however, it is not the same, and the similarity stems from a confusion in the *informal* characterization of the classificatory concept of confirmation. This confusion is cleared up in the preface to the new edition of his book,<sup>202</sup> where Carnap dis-

tinguishes two distinct triples of concepts, neither of which allows Popper's paradox.

#### *firmness*

1.  $h$  is firm on  $e$
2.  $h$  on  $e$  is firmer than  $h'$  on  $e'$
3. the degree of firmness of  $h$  on  $e$  is  $u$ .

#### *increase in firmness*

- $h$  is made firmer by  $i$
- $h$  is made firmer by  $i$  than  $h'$  is by  $i'$
- the increase in firmness of  $h$  by  $i$  is  $u$ .

Each of these concepts admits of a simple and obvious explication in terms of confirmation functions. The paradox observed by Popper only arose through Carnap's mistaken selection of (1) under *increase in firmness* and (2) and (3) under *firmness* as a triple of concepts analogous to "warm," "warmer," and "of such and such a temperature." It is not the case, of course, that  $x$  can be firm on  $z$ , and  $y$  not firm on  $z$ , and  $y$  firmer on  $z$  than is  $x$  on  $z$ . One of the most serious objections to the basic idea of confirmation theory thus turns out to be (as Bar-Hillel observed)<sup>203</sup> a matter of terminological confusion.

## XI.

Confirmation theory does not provide an inductive logic, however, until the problem of establishing a choice among the infinite number of confirmation functions available has been solved. In principle (and winking at some classical curiosities, such as that the degree of confirmation of a universal law will generally be 0) all the problems of inductive logic will be solved by the use of Bayes's theorem—once the required *a priori* probabilities are available. (To choose a confirmation function is of course to accept a set of *a priori* probabilities.) I mentioned that Jeffreys, being of a practical turn of mind, suggested that the problem of *a priori* probabilities could be settled by an international board of scientists. This solution does seem extreme, and a bit authoritarian. But there is one school of thought that avoids the authoritarian overtones of this proposal by going even further in the same direction: it is the subjectivistic or personalistic school of statistical inference.

Although one of the founders of this approach to inductive logic was a philosopher, F. P. Ramsey,<sup>204</sup>

<sup>200</sup> Popper, *The Logic of Scientific Discovery*, p. 390f.

<sup>201</sup> *Ibid.*, p. 391.

<sup>202</sup> Carnap, *The Logical Foundations of Probability*, p. xvi.

<sup>203</sup> Bar-Hillel, "Comments on 'Degree of Confirmation' by Professor K. R. Popper."

<sup>204</sup> Ramsey, *The Foundations of Mathematics*.

it is now mainly defended by mathematical statisticians, such as L. J. Savage,<sup>205</sup> and B. de Finetti.<sup>206</sup> It is perhaps this school more than any other that has been making great strides toward a practical inductive logic in the past few years. These strides may have been made across thin philosophical ice, but they have nonetheless opened up a whole new approach to inductive logic. According to this theory, probability statements are regarded as being essentially subjective, in the sense that any statement may be legitimately assigned any degree of probability by any person. There is, it must quickly be added, also a logical element: the probability statements that express a given person's beliefs must be *coherent*—they must satisfy the axioms of the probability calculus. Thus while any particular statement may be assigned any degree of probability, one assignment will impose restrictions on the freedom with which other assignments may be made. If I assign the value  $p$  to the probability of a statement  $S$ , I am in rationality bound to assign the value  $1-p$  to its negation; and I cannot assign a value greater than  $p$  to the conjunction of  $S$  with another statement.

The most important things to come out of the study of subjective or personalistic probability concern the invariance of posterior probabilities for wide variations in prior probabilities. The earliest of these results were established by de Finetti some time ago. He showed, for example, that though two people may start with widely different degrees of belief concerning the likelihood that an  $A$  kind of thing will be a  $B$  kind of thing, it will nevertheless generally be the case that after the observation of a large number of  $A$  things by both of them, the principles of coherence will force them to have nearly the same degrees of belief concerning the likelihood of a subsequent  $A$  being a  $B$ .<sup>207</sup>

More recently W. Edwards, H. Lindemann, and L. J. Savage have published a long article on (subjectivistic) Bayesian inference in which they discuss in detail some of the circumstances under which posterior distributions of belief are almost entirely determined by the evidence, and are very nearly independent of the initial subjective opinions with which one starts.<sup>208</sup> This important article

does not require more mathematical background than most philosophers who are concerned with inductive logic and confirmation theory have. There is also an accessible discussion of the subjectivistic approach by Savage and holders of frequency or classical views of probability,<sup>209</sup> and there is a book of readings that includes the contribution of Savage just mentioned and the paper by de Finetti as well as papers by other writers.<sup>210</sup>

The fundamental theorem, on this view, is of course Bayes's theorem; there is no problem in finding prior probabilities to feed into it, because these prior probabilities merely reflect our initial opinions made coherent. But the theory has been criticized on the following grounds.<sup>211</sup> There is perfect logical symmetry between premisses and conclusions of the inductive argument with respect to the assignment of probabilities. There is thus no logical reason why I cannot decide in advance what conclusions I wish to believe (and to what degree), and then, in the light of the evidence, adjust my so-called initial probabilities so that, given the evidence that I am in fact given, I shall be "required" by conditions of coherence to attribute just that degree of probability to the conclusions that I began by stipulating that I *would* attribute to them. The only answer to this objection is that, in point of fact, people do not adjust their "initial" probabilities in the light of experience in order to arrive at the posterior probabilities they want to arrive at. Although this is possible *logically*, the subjectivist might say, why should the inductive logician go that far out of his way looking for trouble? This attitude amounts to a denial that the relation between evidence and inductive conclusion is *logical* at all; it is reminiscent of the extreme sociological-anthropological views mentioned earlier: inductive inference proceeds in such and such a way because what we call inductive inferences are inferences that proceed that way. This denial *may* be cogent; it may be that it is impossible on strictly logical grounds to set any limitations to people's degrees of belief. But so far the only *evidence* that this is the case is the fact that as yet no general theory which would impose such logical restrictions has gained anything like universal acceptance.

<sup>205</sup> Savage, *Foundations of Statistics*.

<sup>206</sup> Finetti, "Foresight: Its Logical Laws, Its Subjective Sources."

<sup>207</sup> Finetti, *op. cit.*

<sup>208</sup> Edwards, Lindemann, and Savage, "Bayesian Statistical Inference for Psychological Research."

<sup>209</sup> Savages and others, *The Foundation of Statistical Inference*.

<sup>210</sup> Kyburg and Smokler, *Readings in Subjective Probability*.

<sup>211</sup> Kyburg, *Probability and the Logic of Rational Belief*, and "Comments on Braithwaite's Paper."

## XII.

A fairly recent development in inductive logic has been the introduction of *values* into the evaluation of the evidence bearing on a hypothesis. This approach has been stimulated by developments in statistical decision theory and game theory. These disciplines have studied the characteristics of policies for the "acceptance" or "rejection" of hypotheses. For example, let us consider the choice between two hypotheses:  $H_1$ , which asserts that the proportion of  $A$ 's that are  $B$ 's is  $p_1$ , and  $H_2$ , which asserts that the proportion is  $p_2$ . Let us draw a sample of six  $A$ 's to serve as evidence. (Decision theory is generally concerned with small samples.) Taking only the statistical make-up of the sample as relevant (disregarding the order of  $B$ 's and non- $B$ 's), there are seven possible ratios of  $B$ 's to  $A$ 's in the sample, ranging from 0/6 to 6/6. A *decision rule* tells us when to accept  $H_1$  and when to accept  $H_2$ . One rule would tell us, for example, to accept  $H_1$  if the proportion of  $B$ 's in the sample is 3/6, 5/6, or 6/6, and to accept  $H_2$  otherwise. Obviously some rules are more sensible than others. Since, it must be remembered, we are using "accept" in an uncritically Pickwickian sense here, we cannot evaluate rules without knowing something about the costs of various possible mistakes and the benefits of the various ways of being right. Suppose that we gain  $r_1$  units if we correctly accept  $H_1$ , and  $r_2$  units if we correctly accept  $H_2$ , and that the losses entailed by choosing  $H_1$  when  $H_2$  is true are  $l_2$  and those resulting from the erroneous acceptance of  $H_2$  are  $l_1$ .

If we have *a priori* probabilities for  $H_1$  and  $H_2$ —whether these are subjective, objective, and logical, or objective and derived from some general acceptable theory—then we can compute the best rule to follow by means of the classical theory of mathematical expectation. Let  $s$  stand for the experimental evidence. Then the computation of  $c(s, H_1)$  and of  $c(s, H_2)$  offer no problems in principle, whether or not we have *a priori* probabilities for  $H_1$  and  $H_2$ . Given these *a priori* probabilities, we can compute:

$$c(s) = c(s, H_1) + c(s, H_2) = c(H_1)c(s, H_1) + c(H_2)c(s, H_2)$$

and thus

$$c(H_1, s) = \frac{c(H_1)c(s, H_1)}{c(s)} \quad c(H_2, s) = \frac{c(H_2)c(s, H_2)}{c(s)}$$

to maximize our expectations, we need merely compare

$$r_1c(H_1, s) - l_2c(H_2, s) \quad \text{and} \quad r_2c(H_2, s) - l_1c(H_1, s).$$

This is the way subjectivists and confirmation theorists would handle the problem. Nothing is added by the decision-theorist's frame work except the curious terminology of "accepting" and "rejecting" hypotheses where a terminology concerning actions would seem more natural.

But if we do *not* have the *a priori* probabilities  $c(H_1)$  and  $c(H_2)$ , the situation is different. We can no longer compare mathematical expectations. We must look instead at a  $2 \times n$  table, where  $n$  is the number of possible rules (here  $n = 128$ , the number of ways of dividing the seven possible samples in two groups).

	$H_1$	$H_2$
rule 1	$a_{11}$	$a_{12}$
rule 2	$2a_{21}$	$a_{22}$
.....	.....	.....
rule $n$	$a_{n1}$	$a_{n2}$

The entries in the table represent conditional mathematical expectations under the alternative hypotheses. Thus  $a_{11} = r_1 \times$  (the probability of correctly choosing  $H_1$ , under rule 1, when  $H_1$  is true)  $- l_1 \times$  (the probability of incorrectly choosing  $H_2$  when  $H_1$  is true). The classical minimax policy directs us to choose among the rules by first picking out the minimum entry in each row (the *value* of the rule under the least favorable hypothesis) and then choosing the rule (of the  $n$  rules) which maximizes this minimum value.

Braithwaite,<sup>212</sup> on whose lucid presentation I have modelled mine, has pointed out, following D. V. Lindley,<sup>213</sup> that if we regard the entries  $a_{ij}$  as containing a factor representing the *a priori* probability of the hypothesis  $j$ , then the non-minimax policy of choosing the rule  $k$  for which the sum  $a_{k1} + a_{k2}$  is a maximum will correspond precisely to the Bayesian solution of the decision problem. We may thus use the same matrix whether we are Bayesians or game theorists, and confine our discussions to the choice of a general over-all policy for choosing among rules.

As I have pointed out before,<sup>214</sup> this does not

<sup>212</sup> Braithwaite, *Scientific Explanation*, and "The Role of Values in Scientific Inference."

<sup>213</sup> Lindley, "Statistical Inference."

<sup>214</sup> Kyburg, "Comments on Braithwaite's Paper."

achieve a reconciliation between the two approaches, but it does focus our attention on the fact that if we do regard the numbers  $a_{ij}$  as containing factors representing the *a priori* probabilities of the hypotheses, then there is no doubt about the fact that we should use a Bayesian (least sum) policy. On the other hand, a perfectly coherent and intelligible inductive logic can be based on the assumption that there is no such thing as the *a priori* probability of a hypothesis.

It should also be mentioned that there is considerable ambiguity involved in the notion of the acceptance of a scientific hypothesis. Although it is possible to maintain, as Churchman does, that to accept a hypothesis *H* and to perform the corresponding action mean exactly the same thing,<sup>15</sup> most philosophers, I think, would argue against this strong identification of belief and action; there is no self-contradiction in saying that I shall act on hypothesis *H*, even though I believe that it is false, provided that there is a chance that it is true, and provided the possible gains outweigh the probable losses. On the other hand, it is clear that there is some connection, in most cases, between accepting a hypothesis and acting as if it were true.

### XIII.

There are a number of problems that have come to the fore in the past fifteen years of work in inductive logic. Although the assessment of these problems as important reflects a personal bias, I think it safe to say that those to whom any problems in inductive logic and scientific method are important will find that the following problems—or some of them—are worthy of serious study.

Central to most theories of inductive logic is the notion of probability. Is it possible to provide a logical reconstruction of probability statements which will gain universal acceptance, and which will restrict the freedom of probability assessments allowed by subjectivistic theories? Or is it possible (what seems very unlikely) to show that no such reconstruction can be provided? Or perhaps even that probability, as it is used by philosophers, is a modal term which does not even obey the rules of the probability calculus? Or, finally, must we regard even the term "probability" as irrelevant to issues in the problem of inductive confirmation, and speak instead of "degree of factual support," "degree of corroboration," and so on?

There is a relation, of course, between our point

of view regarding probability and our point of view regarding statistical decisions. If there exist logical *a priori* probabilities, then the only sensible policy is Bayesian. But if these probabilities don't exist, the question as to whether we should introduce subjective probabilities into our statistical reasoning and follow Bayesian procedures, or refrain from introducing them, and follow minimax procedures, may turn out to be a rather deep philosophical question. If we reject even subjectivistic probabilities, then quite clearly we should follow minimax procedures, or some variant of them.

Nearly all of the more or less formal theories of scientific inference that I have discussed above have, at one point or another, run into the Grue-Bleen problem. This problem has many ramifications that haven't been explored yet; and no solution even to the simple cases has gained anything like universal acceptance. Salmon's solution applies only to ostensive predicates, and even then depends on the fact that "looks alike" has a conventional meaning. (I find it easy to imagine creatures to whom all grue things look alike, and who would be horribly surprised if in the year 2000 all grue things suddenly turned bleen.)

There is the continuing problem of elucidating the structure of inductive arguments as they are actually found to be used in scientific inquiry. This can often be done by formalizing the arguments within an ordinary logical symbolism. A particular type of argument which begs for careful analysis and formalization to exhibit its logical structure is statistical inference.

The justification of induction no longer seems to be quite the problem it used to be, because the justification, if any, is now regarded as subsidiary to and dependent on the logic of induction, rather than *vice versa*. In the same way, the search for postulates which, if true, could serve to establish the cogency of scientific reasoning (if only we had grounds for believing them), no longer seems terribly relevant.

I find myself unimpressed by the ordinary-language analysis of inductive arguments, but for those who enjoy this activity the world is full of arguments that have yet to receive this philosophical treatment. I only hope that such analysis removes or takes account of the ambiguities of ordinary language rather than (as we have observed *can* happen) taking advantage of them.

There is one large problem which I have saved for dessert, because it is the problem I found most

<sup>15</sup> Churchman, "Science and Decision Making."

pervasive and at the same time least recognized. It is not a problem which has been dealt with explicitly in any of the papers discussed so far, but I think it underlies much of the argument I have attempted to summarize. It is the problem of whether or not inductive conclusions are ever accepted. The more carefully one attends to the arguments between proponents of the hypothetico-deductive method and inductivists (falsification theorists and confirmation theorists), the more important does this problem appear. Popper regards Carnap as going beyond the pale of true Humean scepticism in saying that hypotheses are "probable"; Popper will *accept* a hypothesis (tentatively, of course), but will never make the assertion that it is probable. But Carnap looks down on Popper for nearly the same reason: it would on his view be altogether unwarranted, and flying in the face of sound Humean sceptical arguments, to *accept* a hypothesis, tentatively or any other way. All we can hope for is to be able to assign some degree of probability to the hypothesis.

What is at issue here is precisely the question of whether we should speak of *accepting* hypotheses; it is the question of whether inductive logic is to have a rule of detachment. (It should be clear that inductive philosophers of a decision-theoretic bent are among those who use the word "accept" in a peculiar sense. They fall into both of the camps being discussed here.)

Carnap is one of the few philosophers to take a very definite stand on this issue, and to recognize it as a fundamentally important one. He says that the great majority of contemporary writers make "one basic mistake": they regard the result of inductive reasoning as the acceptance of a new proposition.<sup>216</sup> On Carnap's view, the result of an induction is the assignment of a degree of confirmation to a new proposition. Now this is all, he would argue, that we need or can possibly want in the way of inductive conclusions. On the basis of these degrees of confirmation, we can define mathematical expectations and make decisions among courses of action. This has also been pointed out by Jeffrey.<sup>217</sup>

Popper, Black, Day, and others, however, have argued that scientists do accept hypotheses, if only

provisionally. Salmon and Hempel have pointed out that there are difficulties involved in supposing they do, and Levi has shown how even on Popper's own scheme the relation between corroboration and acceptance is complicated.<sup>218</sup> The issue was discussed at length at the Wesleyan Conference on Induction.<sup>219</sup> I myself incorporated a rule of detachment in my own system of inductive logic; but it leads to many difficulties, and was indeed the property of the system which enabled Schick to establish its unsuitability. Hempel has expressed some hope that the difficulties involved in a rule of detachment can be overcome.<sup>220</sup>

My own arguments in favor of a rule of detachment are these:

- (1) People do accept statements that are probable enough; they regard them as practically certain. I know that it is only *probable* (in one sense) that I shall find paper in my desk drawer, but in another sense I am *certain* that there is paper there, because I just now put it there. It is unrealistic, a violation of good sense, to demand that all such statements be regarded as only probabilities.
- (2) Among the statements that are "practically certain" are those which report sense experience, and which writers like Carnap must regard as absolutely certain. He must regard such statements as certain in order to have statements relative to which to compute the probabilities of other statements. By incorporating a rule of detachment in our inductive logic, we can avoid this consequence, and we can say that observation statements are only overwhelmingly probable—probable enough to be included directly in our rational corpora, but not so probable as to be incorrigible.
- (3) Carnap deals with the credibility of universal laws through the concept of instance-confirmation. The engineer who uses the laws of mechanics in constructing a bridge is not betting that the laws of mechanics will hold for all the bridges that could ever be built, but only that they will hold for his particular bridge.<sup>221</sup> The degree of instance confirmation that a law will hold true is taken to be the degree of confirmation of the hypothesis that the law will hold in its next application. But few inductive logicians are altogether happy with Carnap's concept of instance-confirmation, and it is possible that a rule of detachment could lead to a more plausible treatment of laws.

<sup>216</sup> Carnap, "The Aim of Inductive Logic."

<sup>217</sup> Jeffrey, "Valuation and Acceptance of Scientific Hypotheses."

<sup>218</sup> Levi, "Corroboration and Rules of Acceptance."

<sup>219</sup> Kyburg and Nagel, *Induction: Some Current Issues*, "Discussion of Salmon's Paper," and "Rejoinder by Salmon."

<sup>220</sup> Hempel, "Inductive Inconsistencies."

<sup>221</sup> Carnap, *The Logical Foundations of Probability*, p. 572.

These arguments may or may not be persuasive, but I think that one of the real problems of inductive logic is still the search for conditions under which a rule of detachment may safely be employed. There are others of a like mind, and it is perhaps significant that J. W. Tukey, a statistician, has recently written in support of a rule of detachment.<sup>222</sup>

There has been real progress in inductive logic in the past fifteen years; more real progress,

perhaps, than one is inclined to expect to find in a branch of philosophy. But this progress has been bought at a price: we have been left with a number of more difficult problems than those with which we started, problems which will require a lot of solid, careful work for their solution. I hope that the foregoing review of events of the past fifteen years will serve to make it easier for those who wish to work on these new problems.

Wayne State University

<sup>222</sup> Tukey, "Conclusions versus Decisions."

## BIBLIOGRAPHY

- ABELSON, Raxiel Review of Lerner, *Evidence and Inference in Philosophy and Phenomenological Research*, Vol. 21 (1960-61), pp. 413-14.
- ABRUZZI, Adam "Problems of Inference in the Socio-Physical Sciences," *Journal of Philosophy*, Vol. 51 (1954), pp. 537-49.
- ACHINSTEIN, Peter "From Success to Truth," *Analysis*, Vol. 21 (1960-61), pp. 6-9.
- "The Circularity of a Self-Supporting Inductive Argument," *Analysis*, Vol. 22 (1961-62), pp. 138-41.
- "Circularity and Induction," *Analysis*, Vol. 23 (1962-63), pp. 123-27.
- See Barker and Achinstein.
- ACKERMANN, Robert "Inductive Simplicity," *Philosophy of Science*, Vol. 28 (1961), pp. 152-61.
- "Some Remarks on Kyburg's Modest Proposal," *Philosophical Review*, Vol. 71 (1962), pp. 236-40.
- AGASSI, J. "Corroboration Versus Induction," *British Journal for the Philosophy of Science*, Vol. 9 (1958-59), pp. 311-17.
- "The Role of Corroboration in Popper's Methodology," *Australasian Journal of Philosophy*, Vol. 39 (1961), pp. 82-91.
- ALEXANDER, H. Gavin "General Statements as Rules of Inference," Feigl, *Minnesota Studies in the Philosophy of Science*, Vol. II, pp. 309-329.
- "The Paradoxes of Confirmation," *British Journal for the Philosophy of Science*, Vol. 9 (1958-59), pp. 227-33.
- "The Paradoxes of Confirmation—A Reply to Dr. Agassi," *British Journal for the Philosophy of Science*, Vol. 10 (1959-60), pp. 229-34.
- "Convention, Falsification, and Induction," *Proceedings of the Aristotelian Society, Supplementary Volume 34* (1960), pp. 131-44.
- ANSCOMBE, F. J. "Mr. Kneale on Probability and Induction," *Mind*, Vol. 60 (1951), pp. 299-309.
- ASHBY, W. Ross "Induction, Prediction, and Decision-Making in Cybernetic Systems," Kyburg and Nagel, *Induction, Some Current Issues*, pp. 55-68.
- AYER, A. J. "The Conception of Probability as a Logical Relation," Körner, *The Colston Papers*, Vol. 9, pp. 12-17.
- "On the Probability of Particular Events," *Revue Internationale de Philosophie*, Vol. 15 (1961), pp. 366-375.
- BAR-HILLEL, Yehoshua "A Note on State Descriptions," *Philosophical Studies*, Vol. 2 (1951), pp. 72-5.
- "A Note on Comparative Inductive Logic," *British Journal for the Philosophy of Science*, Vol. 3 (1952-53), pp. 308-10.
- "Comments on 'Degree of Confirmation' by Professor K. R. Popper," *British Journal for the Philosophy of Science*, Vol. 6 (1955-56), pp. 155-57.
- "Further Comments on Probability and Confirmation," *British Journal for the Philosophy of Science*, Vol. 7 (1956-57), pp. 245-48.
- BARKER, Stephen *Induction and Hypothesis: A Study of the Logic of Confirmation* (Ithaca, New York, Cornell University Press, 1957).
- Review of von Wright, *The Logical Problem of Induction in Journal of Philosophy*, Vol. 55 (1958).
- Review of Jeffreys, *Scientific Inference in Philosophical Review*, Vol. 67 (1958), pp. 404-07.

- BARKER, Stephen "Comments on Salmon's 'Vindication of Induction'," Feigl and Maxwell, *Current Issues in the Philosophy of Science* (1961), pp. 257-60.
- "The Role of Simplicity in Explanation," Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 265-74.
- "Rejoinder to Salmon," Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 276-78.
- "On Simplicity in Empirical Hypotheses," *Philosophy of Science*, Vol. 28 (1961), pp. 162-71.
- BARKER, S. and ACHINSTEIN, P. "On the New Riddle of Induction," *Philosophical Review*, Vol. 69 (1960), pp. 511-22.
- BARTLETT, M. S. "Probability in Logic, Mathematics, and Science," *Dialectica*, Vol. 3 (1949), pp. 104-13.
- BARTLEY, W. W. 3<sup>rd</sup>. "A Note on Barker's Discussion of Popper's Theory of Corroboration," *Philosophical Studies*, Vol. 12 (1961), pp. 5-10.
- BENNETT, Jonathan F. "Some Aspects of Probability and Induction (I)," *British Journal for the Philosophy of Science*, Vol. 7 (1956-57), pp. 220-30.
- "Some Aspects of Probability and Induction (II)," *British Journal for the Philosophy of Science*, Vol. 7 (1956-57), pp. 316-22.
- BERENDA, C. W. "On Verifiability, Simplicity, and Equivalence," *Philosophy of Science*, Vol. 19 (1952-53), pp. 70-6.
- BERGMANN, Gustav "Some Comments on Carnap's Logic of Induction," *Philosophy of Science*, Vol. 13 (1946), pp. 71-78.
- BERLYNE, Daniel E. "The Motivation of Inductive Behavior," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 74-92.
- BEVERIDGE, W. I. B. Review of Wisdom, *Foundations of Inference in Natural Science* in *British Journal for the Philosophy of Science*, Vol. 3 (1952-53), pp. 291-93.
- BLACK, Max *Language and Philosophy* (Ithaca, Cornell University Press, 1949).
- "The Justification of Induction," Black, *Language and Philosophy*, pp. 59-88.
- (Ed.) *Philosophical Analysis* (Englewood Cliffs, N.J., Prentice-Hall, 1963) (first ed. 1950).
- *Problems of Analysis* (Ithaca, Cornell University Press, 1954).
- "Pragmatic Justifications of Induction," Black, *Problems of Analysis*, pp. 157-90.
- "Inductive Support of Inductive Rules," Black, *Problems of Analysis*, pp. 191-208.
- "How Difficult Might Induction Be?" Black, *Problems of Analysis*, pp. 209-25.
- *Models and Metaphors* (Ithaca, Cornell University Press, 1962).
- "Self-Supporting Inductive Arguments," *Journal of Philosophy*, Vol. 55 (1958), pp. 718-725. [Reprinted in Black, *Models and Metaphors*].
- "Can Induction be Vindicated?" *Philosophical Studies*, Vol. 10 (1959), pp. 5-16 [Reprinted in Black, *Models and Metaphors*].
- "Induction and Probability," Klibanaky, *Philosophy in Mid-Century*, pp. 154-63.
- "Self-Support and Circularity: A Reply to Mr. Achinstein," *Analysis*, Vol. 23 (1962-63), pp. 43-4.
- "Comments on Salmon's Paper," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 42-4.
- BLOM, Siri "Concerning a Controversy on the Meaning of 'Probability'," *Theoria*, Vol. 21 (1955), pp. 65-98.
- BROAD, C. D. Review of Kneale, *Probability and Induction in Mind*, Vol. 59 (1950), pp. 94-115.
- BRAITHWAITE, R. B. *Scientific Explanation* (Cambridge University Press, 1953).
- "On Unknown Probabilities," Körner, *The Colston Papers*, Vol. 9, pp. 3-11.
- "The Role of Values in Scientific Inference," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 180-93.
- BRODBECK, May "An Analytic Principle of Induction?" *Journal of Philosophy*, Vol. 49 (1952), pp. 747-50.
- BRONOWSKI, Jacob *The Common Sense of Science* (Cambridge, Harvard University Press, 1953).
- *Science and Human Values* (New York, J. Messner, 1956).
- "The Scandal of Philosophy," *British Journal for the Philosophy of Science*, Vol. 8 (1957-58), pp. 329-34.
- BROWN, G. Spencer *Probability and Scientific Inference* (London, Longman's, Green and Company, 1957).
- "Randomness," *Proceedings of the Aristotelian Society, Supplementary Volume 31* (1957), pp. 145-50.
- BUCHDAHL, Gerd "Convention, Falsification, and Induction," *Proceedings of the Aristotelian Society, Supplementary Volume 34* (1960), pp. 113-30.
- "Induction and Scientific Method," *Mind*, Vol. 60 (1951), pp. 16-34.
- BUNGE, Mario *Metascientific Queries* (Springfield, Illinois, Chas. C. Thomas, 1959).
- "The Place of Induction in Science," *Philosophy of Science*, Vol. 27, (1960), pp. 262-70.
- "The Weight of Simplicity in the Construction and Assaying of Scientific Theories," *Philosophy of Science*, Vol. 28 (1961), pp. 120-49.
- BURKS, Arthur W. "Reichenbach's Theory of Probability and Induction," *Review of Metaphysics*, Vol. 4 (1951), pp. 377-93.
- Review of Carnap, *Logical Foundations of Probability* in *Journal of Philosophy*, Vol. 48 (1951), pp. 524-35.
- "The Presupposition Theory of Induction," *Philosophy of Science*, Vol. 20 (1953), pp. 177-97.

- BURKS, Arthur W. "Justification in Science," White, *Academic Freedom, Logic, and Religion*, pp. 109-25.
- Review of Carnap, *Continuum of Inductive Methods* in *Journal of Philosophy*, Vol. 50 (1953), pp. 731-34.
- "On the Presuppositions of Induction," *Review of Metaphysics*, Vol. 8 (1954-55), pp. 574-611.
- CAMPBELL, Keith "One Form of Scepticism about Induction," *Analysis*, Vol. 23 (1962-63), pp. 80-3.
- CARLSSON, Gösta "Sampling, Probability and Causal Inference," *Theoria*, Vol. 18 (1952), pp. 139-54.
- CARNAP, Rudolf "The Problem of Relations in Inductive Logic," *Philosophical Studies*, Vol. 2 (1951), pp. 75-80.
- "Inductive Logic and Science," *Proceedings of the American Academy of Arts and Sciences*, Vol. 80 (1951-52). [Reprinted as part II of *Statistical and Inductive Probability*.]
- "Meaning Postulates," *Philosophical Studies*, Vol. 3 (1952), pp. 65-73.
- *The Continuum of Inductive Methods* (Chicago, University of Chicago Press, 1952).
- "On the Comparative Concept of Confirmation," *British Journal for the Philosophy of Science*, Vol. 3 (1952-53), pp. 311-18.
- "Remarks to Kemeny's Paper," *Philosophy and Phenomenological Research*, Vol. 13 (1952-53), pp. 375-76.
- *Statistical and Inductive Probability* (Brooklyn, N.Y., The Galois Institute, 1955).
- "Remarks on Popper's Note on Content and Degree of Confirmation," *British Journal for the Philosophy of Science*, Vol. 7 (1956-57), pp. 243-44.
- CARNAP, Rudolf and STEGMÜLLER, Wolfgang *Induktive Logik und Wahrscheinlichkeit* (Wein, Springer, 1959).
- CARNAP, Rudolf *The Logical Foundations of Probability* (2nd edition, Chicago, University of Chicago Press, 1962). [1st edition, 1950.]
- "The Aims of Inductive Logic," Nagel, Suppes, and Tarski, *Logic, Methodology, and Philosophy of Science*, pp. 303-18.
- CASTANEDA, Hector "Are Conditionals Principles of Inference?" *Analysis*, Vol. 18 (1957-58), pp. 77-82.
- "On a Proposed Revolution in Logic," *Philosophy of Science*, Vol. 27 (1960), pp. 279-92.
- CAWS, Peter "The Paradox of Induction and the Inductive Wager," *Philosophy and Phenomenological Research*, Vol. 22 (1961-62), pp. 512-20.
- CHATALIAN, George "Probability: Inductive versus Deductive," *Philosophical Studies*, Vol. 3 (1952), pp. 49-56.
- "Induction and the Problem of the External World," *Journal of Philosophy*, Vol. 49 (1952), pp. 601-07.
- CHISHOLM, Roderick "Epistemic Statements and the Ethics of Belief," *Philosophy and Phenomenological Research*, Vol. 16 (1955-56), pp. 447-60.
- "Appear", "Take", "Evident", *Journal of Philosophy*, Vol. 53 (1956), pp. 722-31.
- "Evidence as Justification," *Journal of Philosophy*, Vol. 58 (1961), pp. 739-48.
- CHURCHMAN, C. West "A Critique of Scientific Critiques," *Review of Metaphysics*, Vol. 7 (1953-54), pp. 89-97.
- "A Pragmatic Theory of Induction," Frank, *The Validation of Scientific Theories*, pp. 18-24.
- "Science and Decision Making," *Philosophy of Science*, Vol. 23 (1956), pp. 247-49.
- *Prediction and Optimal Decision* (Englewood Cliffs, N.J., Prentice-Hall, 1962).
- COBURN, Robert C. "Braithwaite's Inductive Justification of Induction," *Philosophy of Science*, Vol. 28 (1961), pp. 65-71.
- COOLEY, John C. "Toulmin's Revolution in Logic," *Journal of Philosophy*, Vol. 56 (1959), pp. 297-319.
- COPELAND, Arthur H. "Statistical Induction and the Foundations of Probability," *Theoria*, Vol. 28 (1962), pp. 27-44 and 87-109.
- CRAIG, William "Replacement of Auxiliary Expressions," *Philosophical Review*, Vol. 65 (1956), pp. 38-55.
- DAHLSTRÖM, Edmund "Some Aspects of Quantitative Verification in Sociology," *Theoria*, Vol. 17, (1951), pp. 27-38.
- DANTZIG, D. van "Carnap's Foundation of Probability Theory," *Synthese*, Vol. 8 (1953), pp. 459-70.
- DARLINGTON, Jared "On the Confirmation of Laws," *Philosophy of Science*, Vol. 26 (1959), pp. 14-24.
- "Reply to Linhart," *Philosophy of Science*, Vol. 26 (1959), p. 363.
- DAS, R. "Induction and Non-Instantial Hypothesis," *British Journal of the Philosophy of Science*, Vol. 8 (1957-58), pp. 317-25.
- DAY, John P. Review of von Wright, *Logical Problem of Induction in Philosophy*, Vol. 35 (1960), pp. 77-80.
- *Inductive Probability* (New York, Humanities Press, 1961).
- DUCASSE, C. J. "Deductive Probability Arguments," *Philosophical Studies*, Vol. 4 (1953), pp. 29-31.
- EDWARDS, Ward, LINDMAN, Harold and SAVAGE, Leonard J. "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, Vol. 70 (1963), pp. 193-242.
- EWING, A. C. "Causality and Induction," *Philosophy and Phenomenological Research*, Vol. 12 (1951-52), pp. 465-85.
- FEIBLEMAN, J. K. "On the Theory of Induction," *Philosophy and Phenomenological Research*, Vol. 14 (1953-54), pp. 332-42.

- FEIBLEMAN, J. K. "The Logical Structure of the Scientific Method," *Dialectica*, Vol. 13 (1959), pp. 208-25.
- FEIGL, Herbert "The Logical Character of the Principle of Induction," *Philosophy of Science*, Vol. 1 (1934).
- "De Principiis Non Disputandum. . . ?" *Black, Philosophical Analysis*, pp. 113-47.
- "Confirmability and Confirmation," *Revue Internationale de Philosophie*, Vol. 5 (1951).
- "Scientific Method without Metaphysical Pre-suppositions," *Philosophical Studies*, Vol. 5 (1954), pp. 17-29.
- *Minnesota Studies in the Philosophy of Science*, Vol. II, *Concepts, Theories, and the Mind-Body Problem* (Minneapolis, University of Minnesota Press, 1958).
- "On the Vindication of Induction," *Philosophy of Science*, Vol. 28 (1961), pp. 212-16.
- *Minnesota Studies in the Philosophy of Science*, Vol. III, *Scientific Explanation, Space, and Time* (Minneapolis, University of Minnesota Press, 1962).
- FEIGL, Herbert, and MAXWELL, Grover, eds. *Current Issues in the Philosophy of Science* (New York, Holt, Rinehart and Winston, 1961).
- FEUER, Lewis S. "The Principle of Simplicity," *Philosophy of Science*, Vol. 24 (1957), pp. 109-22.
- FEYERABEND, Paul K. "Comments on Hanson's 'Is there a logic of discovery?'," Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 35-9.
- FINCH, Henry A. "Validity Rules for Proportionally Quantified Syllogisms," *Philosophy of Science*, Vol. 24 (1957), pp. 1-18.
- "Confirming Power of Observations Metricized for Decisions Among Hypotheses," *Philosophy of Science*, Vol. 27 (1960), pp. 293-307 (Part I) and pp. 391-404 (Part II).
- FINETTI, Bruno de "Prévision: ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, Vol. 7 (1937), pp. 1-68.
- "Foundations of Probability," Klibansky, *Philosophy in Mid-century*, Pt. I *Logic* (Firenze, 1958), pp. 140-47.
- "Foresight: Its Logical Laws, Its Subjective Sources," Kyburg and Smokler, *Studies in Subjective Probability* (1964). [Translation of "La prévision: ses lois logiques, ses sources subjectives."]
- "La vrai et la Probable," *Dialectica*, Vol. 3 (1949).
- "La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista," *Induzione e Statistica* (Rome, Istituto Matematico Dell' University, 1963).
- FISHER, Ronald A. *Statistical Methods and Scientific Inference* (Edinburgh, Oliver and Boyd, 1959). [First edition, 1956.]
- FISK, Milton "Falsifiability and Corroboration," *Philosophical Studies* (Dublin), Vol. 9 (1959), pp. 49-65.
- Review of Day, *Inductive Probability in Philosophical Studies* (Dublin), Vol. 11 (1961-62).
- FITCH, Frederick B. "Justification in Science," White, *Academic Freedom, Logic and Religion*, pp. 96-107.
- FLEW, Anthony Review of Brown, *Probability and Scientific Inference in Philosophy*, Vol. 9 (1959), pp. 380-81.
- FRANK, Philipp G. (ed.) *The Validation of Scientific Theories*, (Beacon Press, Boston, 1954).
- "The Variety of Reasons for the Acceptance of Scientific Theories," Frank, *The Validation of Scientific Theories*, pp. 3-18.
- *Philosophy of Science: The Link Between Science and Philosophy* (Englewood Cliffs, N.Y., Prentice-Hall, 1957).
- FREUDENTHAL, Hans "Abus philosophiques de la statistique," *Revue de Métaphysique et de Morale*, Vol. 67 (1962), pp. 127-246.
- FREUND, John "On the Problem of Confirmation," *Methodos* 3 (1951), pp. 33-42.
- "On the Confirmation of Scientific Theories," *Philosophy of Science*, Vol. 17 (1950), pp. 87-94.
- FRITZ, Charles A., Jr. "What is Induction?" *Journal of Philosophy*, Vol. 57 (1960), pp. 126-38.
- GEORGESCU-ROEGEN, Nicholas "The End of the Probability Syllogism?" *Philosophical Studies*, Vol. 5 (1954), pp. 31-2.
- GIBBONS, P. C. "On the Severity of Tests," *Australasian Journal of Philosophy*, Vol. 40 (1952), pp. 79-82.
- GOOD, I. J. *Probability and the Weighing of Evidence*, (London, C. Griffin, 1950).
- "Rational Decisions," *Journal of the Royal Statistical Society*, Series B, Vol. 14 (1952), pp. 107-14.
- "Kinds of Probability," *Science*, Vol. 126 (1959), pp. 443-47.
- "The Paradox of Confirmation," *British Journal for the Philosophy of Science*, Vol. 11 (1960-61), pp. 145-49.
- "The Paradox of Confirmation (II)," *British Journal for the Philosophy of Science*, Vol. 12 (1961-62), pp. 63-4.
- "Subjective Probability as a Measure of a Non-Measurable Set," Nagel, Suppes, and Tarski, *Logic, Methodology and Philosophy of Science*, pp. 319-29.
- GOODMAN, Nelson "New Notes on Simplicity," *Journal of Symbolic Logic*, Vol. 17 (1952), pp. 189-91.
- "Axiomatic Measurement of Simplicity," *Journal of Philosophy*, Vol. 52 (1955), pp. 709-22.
- *Fact, Fiction, and Forecast* (Cambridge, Harvard University Press, 1955).
- "Recent Developments in the Theory of Simplicity," *Philosophy and Phenomenological Research*, Vol. 19 (1958-59), pp. 429-46.
- "The Test of Simplicity," *Science*, Vol. 128 (1958), pp. 1054-69.

- GOODMAN, Nelson "Positionality and Pictures," *Philosophical Review*, Vol. 69 (1960), pp. 523-25.
- — "Safety, Strength, Simplicity," *Philosophy of Science*, Vol. 28 (1961), pp. 150-51.
- GROSS, Edward "Toward a Rationale for Science," *Journal of Philosophy*, Vol. 54 (1957), pp. 829-38.
- HAMBLIN, C. L. "The Modal 'Probably,'" *Mind*, Vol. 68 (1959), pp. 234-40.
- HAMLYN, D. W. Review of Harrod, *Foundations of Inductive Logic in Philosophy*, Vol. 33 (1958), pp. 369-70.
- HANSON, Norwood Russell *Patterns of Discovery* (Cambridge University Press, 1958).
- — "The Logic of Discovery," *Journal of Philosophy*, Vol. 55 (1958), pp. 1073-89.
- — "More on 'The Logic of Discovery,'" *Journal of Philosophy*, Vol. 57 (1960), pp. 182-86.
- — "Is There a Logic of Scientific Discovery?" *Australasian Journal of Philosophy*, Vol. 38 (1960), pp. 91-100.
- — "Good Inductive Reasons," *The Philosophical Quarterly*, Vol. 11 (1961), pp. 123-34.
- — "Is There a Logic of Discovery?" Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 20-35.
- HARRÉ, R. "Dissolving the 'Problem' of Induction," *Philosophy*, Vol. 32 (1957), pp. 58-64.
- — "Simplicity as a Criterion of Induction," *Philosophy*, Vol. 34 (1959), pp. 229-34.
- — Review of Barker, *Induction and Hypothesis in Mind*, Vol. 71 (1962), pp. 412-20.
- HARROD, Roy F. "Induction and Probability," *Philosophy*, Vol. 26 (1951), pp. 37-52.
- — "The General Structure of Inductive Argument," *Proceedings of the Aristotelian Society*, Vol. 52 (1951-52), pp. 41-56.
- — *The Foundations of Inductive Logic* (New York, Harcourt Brace and Company, 1956).
- — "New Argument for Induction: Reply to Professor Popper," *British Journal for the Philosophy of Science*, Vol. 10 (1959-60), pp. 309-12.
- HARSANYI, John C. "Popper's Improbability Criterion for the Choice of Scientific Hypotheses," *Philosophy*, Vol. 35 (1960), pp. 332-40.
- HAY, William H. "Professor Carnap and Probability," *Philosophy of Science*, Vol. 19 (1952), pp. 170-77.
- — Review of Carnap, *Continuum of Inductive Methods in Philosophical Review*, Vol. 62 (1953), pp. 468-72.
- — Review of von Wright, *Treatise on Probability and Induction in Journal of Philosophy*, Vol. 50 (1953), pp. 782-88.
- HAYEK, F. "Degrees of Explanation," *British Journal for the Philosophy of Science*, Vol. 6 (1955-56), pp. 209-25.
- HEMPEL, Carl G. "Inductive Inconsistencies," *Synthese*, Vol. 12 (1960), pp. 439-69.
- — "Empirical Statements and Falsifiability," *Philosophy*, Vol. 33 (1958), pp. 342-48.
- — "Deductive-Nomological vs. Statistical Explanation," Feigl, *Minnesota Studies in the Philosophy of Science*, Vol. III, pp. 98-169.
- HESSE, Mary B. Review of Jeffreys, *Scientific Inference in Philosophy*, Vol. 34 (1959), pp. 66-8.
- HILLMAN, Donald "The Measurement of Simplicity," *Philosophy of Science*, Vol. 29 (1962), pp. 225-52.
- HIRST, P. H. Review of Barker, *Induction and Hypothesis in Philosophical Quarterly*, Vol. 10 (1960), pp. 375-76.
- HJORTH, Sune "The Meanings of Probability Statements," *Theoria*, Vol. 25 (1959), pp. 27-30.
- HOFSTADTER, Albert "Quine's View of Knowledge: The Myth of the Whole," *Journal of Philosophy*, pp. 397-417.
- HOSIASSON-LINDENBAUM, Janina "On Confirmation," *Journal of Symbolic Logic*, Vol. 5 (1940), pp. 133-48.
- HUTTEN, E. H. "Probability-Sentences," *Mind*, Vol. 61 (1952), pp. 38-56.
- — Review of Day, *Inductive Probability in Mind*, Vol. 71 (1962), p. 583.
- ISSMAN, S. "Les problèmes de la déduction logique et des inférences inductives," *Revue Internationale de Philosophie*, Vol. 13 (1959), pp. 132-34.
- JEFFREY, Richard C. "Comments on Leblanc Paper," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 18-21.
- — "Valuation and Acceptance of Scientific Hypotheses," *Philosophy of Science*, Vol. 23 (1956), pp. 237-46.
- JEFFREYS, Harold "Bertrand Russell on Probability," *Mind*, Vol. 59 (1950), pp. 313-19.
- — Review of von Wright, *A Treatise on Induction and Probability in British Journal for the Philosophy of Science*, Vol. 3 (1952-53), pp. 276-77.
- — "The Present Position in Probability Theory," *British Journal for the Philosophy of Science*, Vol. 5 (1954-55), pp. 275-89.
- — *Scientific Inference* [second edition] (Cambridge University Press, 1957). [First edition, 1931.]
- JUHOS, Bela "Deduktion, Induktion und Wahrscheinlichkeit," *Methodos*, Vol. 6 (1954), pp. 259-76.
- KADING, Daniel "Concerning Mr. Feigl's 'Vindication' of Induction," *Philosophy of Science*, Vol. 27 (1960), pp. 405-07.
- KATZ, Jerrold J. *The Problem of Induction and Its Solution* (Chicago, University of Chicago Press, 1962).
- KEENE, G. B. "Randomness," *Proceedings of the Aristotelian Society, Supplementary Volume 31* (1957), pp. 151-60.
- — "Mill's Method of Hypothesis," *Filosofia, Supplementary Volume 13* (1962), pp. 595-98.
- KEMENY, John G. Review of Carnap, *Logical Foundations of Probability in Journal of Symbolic Logic*, Vol. 16 (1951), pp. 205-07.

- KEMENY, John G. Review of Carnap, *Logical Foundations of Probability* in *Review of Metaphysics*, Vol. 5 (1951-52), pp. 145-56.
- "Extension of the Methods of Inductive Logic," *Philosophical Studies*, Vol. 3 (1952), pp. 38-42.
- "A Contribution to Inductive Logic," *Philosophy and Phenomenological Research*, Vol. 13 (1952-53), pp. 371-74.
- "A Logical Measure Function," *Journal of Symbolic Logic*, Vol. 18 (1953), pp. 289-308.
- Review of Wright, *A Treatise on Induction and Probability* in *Philosophical Review*, Vol. 62 (1953), pp. 93-101.
- "The Use of Simplicity in Induction," *Philosophical Review*, Vol. 62 (1953), pp. 391-408.
- "Two Measures of Complexity," *Journal of Philosophy*, Vol. 52 (1955), pp. 722-33.
- "Fair Bets and Inductive Probabilities," *Journal of Symbolic Logic*, Vol. 20 (1955), pp. 263-73.
- *A Philosopher Looks at Science* (Princeton, N.J., D. VanNostrand, 1959).
- KEMENY, John G. and OPPENHEIM, Paul "Degree of Factual Support," *Philosophy of Science* (1952), pp. 307-24.
- KESSEN, William, "Comments on Berlyne's Paper," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 94-97.
- KHATCHADOURIAN, Haig "Some Metaphysical Presuppositions of Science," *Philosophy of Science*, Vol. 22 (1955).
- KING-FARLOW, John "Toulmin's Analysis of Probability," *Theoria*, Vol. 29 (1963), pp. 12-26.
- KLIBANSKY, Raymond *Philosophy in Mid-Century*, Pt. I: *Logic* (Firenze, 1958).
- KNEALE, William *Probability and Induction* (Oxford University Press, 1949).
- "Probability and Induction," *Mind*, Vol. 60 (1951), pp. 310-17.
- "Some Aspects of Probability and Induction: A Reply to Mr. Bennett," *British Journal for the Philosophy of Science*, Vol. 8 (1957-58), pp. 57-63.
- KNEEBONE, George T. "Frequency Theory of Induction," *Proceedings of the Aristotelian Society*, Vol. 50 (1949-50), pp. 27-42.
- "Induction and Probability," *Philosophy*, Vol. 26 (1951), pp. 261-62.
- KÖRNER, Stephen, ed. *The Colston Papers 9, Observation and Interpretation: A Symposium of Philosophers and Scientists* (London, 1957).
- KOSLOW, Arnold Review of Jeffreys, *Scientific Inference* in *Journal of Philosophy*, Vol. 57 (1960), pp. 384-91.
- KUHN, Thomas S. "Historical Structure of Scientific Discovery," *Science*, Vol. 136 (1962), pp. 760-64.
- KYBURG, Henry E., Jr. "The Justification of Induction," *Journal of Philosophy*, Vol. 53 (1956), pp. 392-400.
- "R. B. Braithwaite on Probability and Induction," *British Journal for the Philosophy of Science*, Vol. 9 (1958-59), pp. 203-20.
- "Demonstrative Induction," *Philosophy and Phenomenological Research*, Vol. 21 (1960-61), pp. 80-92.
- "Probability and Rationality," *The Philosophical Quarterly*, Vol. 11 (1961), pp. 3-10.
- "A Modest Proposal Concerning Simplicity," *The Philosophical Review*, Vol. 70 (1961), pp. 390-95.
- *Probability and the Logic of Rational Belief* (Middletown, Wesleyan University Press, 1961).
- Review of Pap, *Introduction to the Philosophy of Science* in *Journal of Philosophy*, Vol. 60 (1963), pp. 358-62.
- Review of Carnap, *Logical Foundations of Probability* (second edition) in *Journal of Philosophy*, Vol. 60 (1963), pp. 362-64.
- "Comments on Braithwaite's Paper," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 196-99.
- "Probability and Randomness," *Theoria*, Vol. 29 (1963), pp. 27-55.
- "A Further Note on Rationality and Consistency," *Journal of Philosophy*, Vol. 60 (1963), pp. 463-65.
- KYBURG and NAGEL (eds.) *Induction: Some Current Issues* (Middletown, Conn., Wesleyan University Press, 1963).
- KYBURG, Henry and SMOKLER, Howard (eds.) *Readings in Subjective Probability* (New York, John Wiley and Sons, 1964).
- LAER, P. Henry Van *Philosophy of Science* (Pittsburgh, Duquesne Studies, Duquesne University Press, 1956).
- LEBLANC, Hugues "Évidence logique et degré de confirmation," *Revue Philosophique de Louvain*, Vol. 52 (1954), pp. 619-25.
- "Two Probability Concepts," *Journal of Philosophy*, Vol. 53 (1956), pp. 679-88.
- "On Chances and Estimated Chances of Being True," *Revue Philosophique de Louvain*, Vol. 57 (1959), pp. 225-39.
- "On So-Called Degrees of Confirmation," *British Journal for the Philosophy of Science*, Vol. 10 (1959-60), pp. 312-15.
- "Probabilities as Truth Value Estimates," *Philosophy of Science*, Vol. 28 (1961), pp. 414-17.
- "A New Interpretation of  $c(h, e)$ ," *Philosophy and Phenomenological Research*, Vol. 21 (1960-61), pp. 373-76.
- "The Problem of the Confirmation of Laws," *Philosophical Studies*, Vol. 12 (1961), pp. 81-4.
- "Statistical and Inductive Probabilities," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 3-16.
- *Statistical and Inductive Probabilities* (Englewood Cliffs, N.J., Prentice-Hall, 1962).
- "That Positive Instances Are No Help," *Journal of Philosophy*, Vol. 60 (1963), pp. 453-62.

- LEBLANC, Hugues "A Revised Version of Goodman's Paradox on Confirmation," *Philosophical Studies*, Vol. 14 (1963), pp. 49-51.
- LEHMAN, R. Sherman "On Confirmation and Rational Betting," *Journal of Symbolic Logic*, Vol. 20 (1955), pp. 251-52.
- LENZ, John W. "Carnap on Defining 'Degrees of Confirmation'," *Philosophy of Science*, Vol. 23 (1956), pp. 230-36.
- "Problems for the Practicalists' Justification of Induction," *Philosophical Studies*, Vol. 9 (1958).
- LERNER, Daniel (ed.) *Evidence and Inference* (Glencoe, Ill., The Free Press, 1959).
- LEVI, Isaac "Must the Scientist Make Value Judgments?" *Journal of Philosophy*, Vol. 57 (1960), pp. 345-57.
- Review of Harrod, *Foundation of Inductive Logic* in *Journal of Philosophy*, Vol. 55 (1958), pp. 209-12.
- "Decision Theory and Confirmation," *Journal of Philosophy*, Vol. 58 (1961), pp. 614-25.
- "On the Seriousness of Mistakes," *Philosophy of Science*, Vol. 29 (1962), pp. 47-65.
- "Corroboration and Rules of Acceptance," *British Journal for the Philosophy of Science*, Vol. 13 (1962-63), pp. 307-13.
- LEVY, Paul "Le fondement du calcul des probabilités," *Revue de Métaphysique*, Vol. 59 (1954), pp. 164-79.
- LEWIS, H. D. (ed.) *Contemporary British Philosophy* (New York, 1956).
- LINDLEY, D. V. "Statistical Inference," *Journal of the Royal Statistical Society*, Series B, Vol. 15 (1953), pp. 30-76.
- LINHART, H. "Darlington's 'On the Confirmation of Laws'," *Philosophy of Science*, Vol. 26 (1949), p. 362.
- LLEWELYN, J. E. "Quantified Inductive Generalizations," *Analysis*, Vol. 22 (1961-62), pp. 134-37.
- MACE, Cecil Alec (ed.) *British Philosophy in Mid-century* (London, Allen and Unwin, 1957).
- MACKIE, J. L. "The Paradox of Confirmation," *British Journal for the Philosophy of Science*, Vol. 13 (1962-63), pp. 265-77.
- McLENDON, Hiram J. "Has Russell Answered Hume?" *Journal of Philosophy*, Vol. 49 (1952), pp. 145-59.
- McNABB, D. G. C. "Hume on Induction," *Revue Internationale de Philosophie*, Vol. 6 (1952), pp. 184-98.
- MADDEN Edward "Aristotle's Treatment of Probability and Signs," *Philosophy of Science*, Vol. 24 (1957), pp. 167-72.
- Review of von Wright, *Logical Problem of Induction in Philosophy and Phenomenological Research*, Vol. 18 (1957-58), pp. 550-51.
- "The Riddle of Induction," *Journal of Philosophy*, Vol. 55 (1958), pp. 705-18.
- *The Structure of Scientific Thought* (Boston, Houghton-Mifflin, 1960).
- MAHALANOBIS, P. C. "The Foundations of Statistics," *Dialectica*, Vol. 8 (1954), pp. 95-111.
- MARGOLIS, Joseph "The Demand for a Justification of Induction," *Synthese*, Vol. 11 (1959), pp. 259-64.
- MARTIN, Richard M. "A Formalization of Inductive Logic," *Journal of Symbolic Logic*, Vol. 23 (1958), pp. 251-56.
- MATSON, W. I. "Against Induction and Empiricism," *Proceedings of the Aristotelian Society*, Vol. 62 (1961-62), pp. 143-58.
- MAXWELL, Grover "An 'Analytic' Vindication of Induction," *Philosophical Studies*, Vol. 12 (1961), pp. 43-45.
- See Feigl and Maxwell.
- MISES, Richard von *Probability, Statistics, and Truth* (Macmillan, New York, 1957). (Translation of the German edition, revised 1951).
- MOORE, Asher "The Principle of Induction," *Journal of Philosophy*, Vol. 49 (1952), pp. 741-47.
- "The Principle of Induction (II): A Rejoinder to Miss Brodbeck," *Journal of Philosophy*, Vol. 49 (1952), pp. 750-58.
- MORGENBESSER, Sidney "A Note on Justification," *Journal of Philosophy*, Vol. 58 (1961), pp. 748-49.
- MUNDLE, C. W. K. "Probability and Scientific Inference," *Philosophy*, Vol. 34 (1959), pp. 150-54.
- NAGEL, Ernest *The Structure of Science* (New York, Harcourt, Brace, and World, 1961).
- NAGEL, SUPPES, and TARSKI (eds.) *Logic, Methodology and Philosophy of Science* (Stanford, Stanford University Press, 1962).
- NAGEL See Kyburg and Nagel.
- NELSON, John O. "The Confirmation of Hypotheses," *Philosophical Review*, Vol. 67 (1958), pp. 95-100.
- "Are Inductive Generalizations Quantifiable?" *Analysis*, Vol. 22 (1961-62), pp. 59-65.
- NEYMAN, Jerzy "Raisonnement inductif ou comportement inductif," *Proceedings of the International Statistical Conference*, Vol. 3 (1947), pp. 423-33.
- *Lectures and Conferences on Mathematical Statistics and Probability*. Graduate School, U.S. Department of Agriculture, Washington (1952) (2nd ed.).
- "The Problem of Inductive Inference," *Communications on Pure and Applied Mathematics*, Vol. 8 (1955), pp. 13-45.
- "Inductive Behavior' as a Basic Concept of Philosophy of Science," *Review of the International Statistical Institute*, Vol. 25 (1957), pp. 7-22.
- NIELSEN, Harry A. "Sampling and the Problem of Induction," *Mind*, Vol. 68 (1959), pp. 474-81.
- NOWAK, Stefan "Some Problems of Causal Interpretation of Statistical Relationships," *Philosophy of Science*, Vol. 27 (1960), pp. 23-38.
- NYMAN, Alf "Induction et Intuition," *Theoria*, Vol. 19 (1953), pp. 21-41.

- OLIVER, James Willard "Deduction and Statistical Syllogism," *Journal of Philosophy*, Vol. 50 (1953), pp. 805-07.
- OLIVER, W. Donald "A Re-Examination of the Problem of Induction," *Journal of Philosophy*, Vol. 49, 1952, pp. 769-80.
- OPPENHEIM See Kemeny.
- O'TOOLE, Edward J. "A Note on Probability," *Philosophical Studies* (Dublin) Vol. 11 (1961-62), pp. 112-27.
- PAKSWER, S. "Information, Entropy, and Inductive Logic," *Philosophy of Science*, Vol. 21 (1954), pp. 254-59.
- PALMIERI, L. E. "Confirmation, Intention, and Language," *Methodos*, Vol. 11 (1959), pp. 33-40.
- "Induction, Deduction, and Certainty," *Methodos*, Vol. 11 (1959), pp. 169-72.
- PAP, Arthur *An Introduction to the Philosophy of Science* (Glencoe, New York, The Free Press of Glencoe, 1962).
- PASSMORE, J. A. "Popper's Account of Scientific Method," *Philosophy*, Vol. 35 (1960), pp. 326-31.
- PETTIJOHN, William C. "Salmon on 'the Short Run'," *Philosophy of Science*, Vol. 23 (1956), p. 149.
- POPPER, Karl R. "Degree of Confirmation," *British Journal for the Philosophy of Science*, Vol. 5 (1954-55), pp. 143-49.
- "Two Autonomous Axiom Systems for the Calculus of Probabilities," *British Journal for the Philosophy of Science*, Vol. 6 (1955-56), pp. 51-7.
- "'Content' and 'Degree of Confirmation': A Reply to Dr. Bar-Hillel," *British Journal for the Philosophy of Science*, Vol. 6 (1955-56), pp. 157-63.
- "Three Views Concerning Human Knowledge," in Lewis, *Contemporary British Philosophy*, pp. 355-88. Reprinted in Popper, *Conjectures and Refutations*, pp. 97-119.
- "Reply to Professor Carnap," *British Journal for the Philosophy of Science*, Vol. 7 (1956-57), pp. 244-45.
- "Adequacy and Consistency: A Second Reply to Dr. Bar-Hillel," *British Journal for the Philosophy of Science*, Vol. 7 (1956-57), pp. 249-56.
- "A Second Note on Degree of Confirmation," *British Journal for the Philosophy of Science*, Vol. 7, 1956-57, pp. 350-53. Reprinted in appendix \*ix, of *The Logic of Scientific Discovery*.
- "Probability Magic or Knowledge Out of Ignorance," *Dialectica*, Vol. 11 (1957), pp. 354-73.
- "The Propensity Interpretation of the Calculus of Probability and the Quantum Theory," *The Colston Papers*, ed. Korner, pp. 65-70.
- "Science: Conjectures and Refutations," *Conjectures and Refutations*, pp. 33-65. First in Mace, (ed.) *British Philosophy in Mid-Century*.
- "A Third Note on Degree of Corroboration or Confirmation," *British Journal for the Philosophy of Science*, Vol. 8 (1957-58), pp. 294-302. Reprinted in appendix \*ix of *The Logic of Scientific Discovery*.
- "The Aim of Science," *Ratio*, Vol. 1 (1957-58), pp. 24-35.
- "On Mr. Harrod's New Argument for Induction," *British Journal for the Philosophy of Science*, Vol. 9 (1958-59), pp. 221-24.
- *The Logic of Scientific Discovery* (London, Hutchinson and Co., 1959). [First published as *Logik der Forschung* in 1934.]
- "The Propensity Interpretation of Probability," *British Journal for the Philosophy of Science*, Vol. 10 (1959-60), pp. 25-42.
- "Probabilistic Independence and Corroboration by Empirical Tests," *British Journal for the Philosophy of Science*, Vol. 10 (1959-60), pp. 315-18.
- "On the Sources of Knowledge and Ignorance," *Conjectures and Refutations*, pp. 23-30. First published in *Proceedings of the British Academy*, Vol. 46 (1960).
- "Truth, Rationality, and the Growth of Scientific Knowledge," *Conjectures and Refutations*, pp. 215-50. [Prepared for Stanford Congress, 1960.]
- "On Carnap's Version of LaPlace's Rule of Succession," *Mind*, Vol. 71, (1962), pp. 69-73.
- "The Demarcation Between Science and Metaphysics," *Conjectures and Refutations*, pp. 253-92. [Prepared for the Schlipp volume on Carnap.]
- *Conjectures and Refutations* (London, Routledge and Kegan Paul, 1963).
- POST, H. R. "Simplicity in Scientific Theories," *British Journal for the Philosophy of Science*, Vol. 11 (1960-61), pp. 32-41.
- PUTNAM, Hilary "A Definition of Degree of Confirmation for Very Rich Languages," *Philosophy of Science*, Vol. 23 (1956), pp. 58-62.
- QUINE, W. V. O. *From a Logical Point of View* (Cambridge, Harvard University Press, 1953).
- "The Scope and Language of Science," *British Journal for the Philosophy of Science*, Vol. 8 (1957-58), pp. 1-17.
- RAMSEY, F. P. *The Foundations of Mathematics* (London, Routledge and Kegan Paul, 1931).
- REICHENBACH, Hans "On the Justification of Induction," *Journal of Philosophy*, Vol. 37 (1940), pp. 97-103.
- *The Theory of Probability* (Berkeley and Los Angeles, University of California Press, 1949).
- RESCHER, Nicholas "Can There Be Random Individuals?" *Analysis*, Vol. 18 (1957-58), pp. 114-17.
- "A Contribution to Modal Logic," *Review of Metaphysics*, Vol. 12 (1958), pp. 186-99.
- "Theory of Evidence," *Philosophy of Science*, Vol. 25 (1958), pp. 83-94.
- "The Problem of a Logical Theory of Belief Statements," *Philosophy of Science*, Vol. 27 (1960), pp. 98-95.
- "Plausible Implication," *Analysis*, Vol. 21 (1960-61), pp. 128-35.

- RESCHER, Nicholas "Non-deductive Rules of Inference and Problems in the Analysis of Inductive Reasoning," *Synthese*, Vol. 13 (1961), pp. 242-51.
- "The Concept of Randomness," *Theoria*, Vol. 27 (1961), pp. 1-11.
- "On the Probability of Non-recurring Events," Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 228-37.
- "Discrete State Systems, Markov Chains, and Problems in the Theory of Scientific Explanation and Prediction," *Philosophy of Science*, Vol. 30, (1963), pp. 325-45.
- RESNICK, Lawrence "Confirmation and Hypothesis," *Philosophy of Science*, Vol. 26 (1959), pp. 25-30.
- ROBBINS, Herbert "A New Approach to a Classical Statistical Decision Problem," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 101-10.
- ROSTHAL, Robert Review of Barker, *Induction and Hypothesis in Philosophy and Phenomenological Research*, Vol. 19 (1958-59), pp. 123-24.
- ROZEBOOM, William W. "Ontological Induction and the Logical Typology of Scientific Variables," *Philosophy of Science*, Vol. 28 (1961), pp. 337-77.
- RUDNER, Richard S. "The Scientist Qua Scientist Makes Value Judgments," *Philosophy of Science*, Vol. 20 (1953), pp. 1-6.
- "Value Judgments in the Acceptance of Theories," Frank, *The Validation of Scientific Theories*, pp. 24-8.
- "Comments on Salmon's 'Vindication of Induction,'" Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 262-64.
- "An Introduction to Simplicity," *Philosophy of Science*, Vol. 28 (1961), pp. 109-19.
- RUSSELL, Bertrand *Human Knowledge: Its Scope and Limits* (New York, Simon Schuster, 1948).
- RUSSELL, L. J. Review of Carnap, *Continuum of Inductive Methods in Philosophy*, Vol. 28 (1953), pp. 272-73.
- RYLE, Gilbert "If, 'So,' and 'Because,'" Black, *Philosophical Analysis*, pp. 302-18.
- "Predicting and Inferring," Körner, *The Colston Papers* 9, pp. 165-70.
- "Comment on Mr. Achinstein's Paper," *Analysis*, Vol. 21 (1960-61), pp. 9-11.
- SALMON, Wesley "The Frequency Interpretation and Antecedent Probabilities," *Philosophical Studies*, Vol. 4 (1953), pp. 44-8.
- "The Uniformity of Nature," *Philosophy and Phenomenological Research*, Vol. 14 (1953-54), pp. 39-48.
- "The Short Run," *Philosophy of Science*, Vol. 22 (1955), pp. 214-21.
- "Regular Rules of Induction," *Philosophical Review*, Vol. 65 (1956), pp. 385-88.
- "Reply to Pettijohn," *Philosophy of Science*, Vol. 23 (1956), pp. 150-51.
- "The Predictive Inference," *Philosophy of Science*, Vol. 24 (1957), pp. 180-90.
- "Should We Attempt to Justify Induction?" *Philosophical Studies*, Vol. 8, (1957), pp. 33-48.
- Review of Barker, *Science and Hypothesis in Philosophical Review*, Vol. 68 (1959).
- "Vindication of Induction," Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 245-57.
- "Rejoinder to Barker," Feigl and Maxwell, *Current Issues in the Philosophy of Science*, pp. 260-62.
- "On Vindicating Induction," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 27-41.
- Review of Day, *Inductive Probability in Philosophical Review*, Vol. 72 (1963), pp. 392-96.
- SAVAGE, Leonard J. *Foundations of Statistics* (New York, John Wiley, 1954).
- "The Foundations of Statistics Reconsidered," *Proceedings of the Fourth [1960] Berkeley Symposium on Mathematics and Probability* (Berkeley, University of California Press, 1961), pp. 575-85.
- "Subjective Probability and Statistical Practice," Savage and others, *Foundations of Statistical Inference*.
- SAVAGE, L. J. and Others *The Foundations of Statistical Inference* (New York, John Wiley and Sons, 1962).
- SCHEFFLER, Israel "A Note on Confirmation," *Philosophical Studies*, Vol. 11 (1960), pp. 21-3.
- "Inductive Inference: a New Approach," *Science*, Vol. 127 (1958), pp. 177-81.
- "A Rejoinder on Confirmation," *Philosophical Studies*, Vol. 12 (1961), pp. 19-20.
- SCHICK, Frederick "Rationality and Consistency," *Journal of Philosophy*, Vol. 60 (1963), pp. 5-19.
- SCHON, Donald "Comments on Mr Hanson's 'The Logic of Discovery,'" *Journal of Philosophy*, Vol. 56 (1959), pp. 500-503.
- SCRIVEN, Michael "The Principle of Inductive Simplicity," *Philosophical Studies*, Vol. 6 (1955), pp. 26-30.
- SELLARS, Wilfrid "Inference and Meaning," *Mind*, Vol. 62 (1953), pp. 313-38.
- SHIMAN, Paul L. "Comments on Ashby Paper," Kyburg and Nagel, *Induction: Some Current Issues*, pp. 68-70.
- SHIMONY, Abner "Braithwaite on Scientific Method," *Review of Metaphysics*, Vol. 7 (1953-54), pp. 644-60.
- "Coherence and the Axioms of Confirmation," *Journal of Symbolic Logic*, Vol. 20 (1955), pp. 1-28.
- SIMON, Herbert A. "Prediction and Hindsight as Confirmatory Evidence," *Philosophy of Science*, Vol. 22 (1955), pp. 227-30.
- SMALL, Kenneth "Professor Goodman's Puzzle," *Philosophical Review*, Vol. 70 (1961), pp. 544-52.
- SMART, J. C. C. "Excogitation and Induction," *Australasian Journal of Philosophy*, Vol. 28 (1950), pp. 191-99.
- SMOKLER, Howard See Kyburg and Smokler.
- SMULLYAN, Arthur "The Concept of Empirical Knowledge," *Philosophical Review*, Vol. 65 (1956), pp. 362-70.

- SMULLYAN, Arthur *Fundamentals of Logic* (Englewood Cliffs, N.J., Prentice-Hall, 1962).
- SOMEZI, Vittorio "Can Induction be Mechanized?" *Methodos*, Vol. 7 (1955), pp. 147-51.
- SPILSBURY, R. J. Review of Churchman, *Theory of Experimental Inference in Mind*, Vol. 59 (1950), pp. 115-16.
- STEGMÜLLER, Wolfgang See Carnap and Stegmüller.
- STOPES-ROE, Harry V. "Recipes and Induction: Ryle v. Achinstein," *Analysis*, Vol. 21 (1960-61), pp. 115-20.
- STOVE, D. C. Review of Harrod, *Foundations of Inductive Logic* in *Australasian Journal of Philosophy*, Vol. 36 (1958), pp. 71-9.
- — "Popperian Confirmation and the Paradox of the Ravens," *Australasian Journal of Philosophy*, Vol. 37 (1959), pp. 149-55.
- — "A Reply to Mr. Watkins," *Australasian Journal of Philosophy*, Vol. 38 (1960), pp. 51-4.
- — Review of Popper, *Logic of Scientific Discovery* in *Australasian Journal of Philosophy*, Vol. 38 (1960), pp. 173-87.
- STRAWSON, Peter F. *Introduction to Logical Theory* (New York, John Wiley, 1952).
- — "On Justifying Induction," *Philosophical Studies*, Vol. 9 (1958), pp. 20-1.
- SUPPES, Patrick "Nelson Goodman on the Concept of Logical Simplicity," *Philosophy of Science*, Vol. 23 (1956), pp. 153-59.
- — "The Philosophical Relevance of Decision Theory," *Journal of Philosophy*, Vol. 58 (1961), pp. 605-14.
- SVENONIUS, Lars "Definability and Simplicity," *Journal of Symbolic Logic*, Vol. 20 (1955), pp. 235-250.
- TOULMIN, Stephen *The Philosophy of Science* (London, Hutchinson's University Library, 1953).
- — *The Uses of Argument* (Cambridge University Press, 1958).
- TUKEY, John W. "Conclusions versus Decisions," *Technometrics*, Vol. 2 (1960), pp. 423-33.
- ULLIAN, Joseph "More on 'Grue' and 'Grue,'" *Philosophical Review*, Vol. 70 (1961), pp. 386-89.
- — "Luck, License, and Lingo," *Journal of Philosophy*, Vol. 58 (1961), pp. 731-38.
- VINCENT, R. H. "A Note on Some Quantitative Theories of Confirmation," *Philosophical Studies*, Vol. 12 (1961), pp. 91-2.
- — "Popper on Qualitative Confirmation and Disconfirmation," *Australasian Journal of Philosophy*, Vol. 40 (1962), pp. 159-66.
- — "The Paradox of Ideal Evidence," *Philosophical Review*, Vol. 71 (1962), pp. 497-503.
- WARNOCK, G. J. Review of Popper, *Logic of Scientific Discovery* in *Mind*, Vol. 69 (1960), pp. 99-101.
- WATKINS, J. W. N. "Between Analytic and Empirical," *Philosophy*, Vol. 32 (1957), pp. 112-31.
- — "A Rejoinder to Professor Hempel's Reply," *Philosophy*, Vol. 33 (1958), pp. 349-55.
- — "Mr. Stove's Blunder's," *Australasian Journal of Philosophy*, Vol. 37 (1959), pp. 240-41.
- — "A Reply to Mr. Stove's Reply," *Australasian Journal of Philosophy*, Vol. 38 (1960), pp. 54-8.
- — "Confirmation without Background Knowledge," *British Journal for the Philosophy of Science*, Vol. 10 (1959-60), pp. 318-20.
- — "Professor Scheffler's Note," *Philosophical Studies*, Vol. 12 (1961), pp. 16-19.
- WHITE, Morton (ed.) *Academic Freedom, Logic, and Religion* (Philadelphia, University of Pennsylvania Press, 1953).
- WHITELY, C. H. Review of Wisdom, *Foundations of Inference in the Natural Sciences in Mind*, Vol. 62 (1953), pp. 113-14.
- WILL, Frederick L. "Generalization and Evidence," Black, *Philosophical Analysis*, pp. 359-86.
- — "Kneale's Theories of Probability and Induction," *Philosophical Review*, Vol. 63 (1954), pp. 19-42.
- — "The Justification of Theories," *Philosophical Review*, Vol. 64 (1955), pp. 370-88.
- — "Justification and Induction," *Philosophical Review*, Vol. 68 (1959), pp. 359-72.
- WILLIAMS, Donald C. *The Ground of Induction* (Cambridge, Harvard University Press, 1947).
- — Review of Reichenbach, *Theory of Probability* in *Philosophy and Phenomenological Research*, Vol. 11 (1950-51), pp. 252-57.
- — "Professor Carnap's Philosophy of Probability," *Philosophy and Phenomenological Research*, Vol. 13 (1952-53), pp. 103-121.
- — "The External World and Mr. Chatalian," *Journal of Philosophy*, Vol. 50 (1953), pp. 13-18.
- WILLIAMS, Donald C. "On the Direct Probability of Inductions," *Mind*, Vol. 62 (1953), pp. 465-83.
- WISDOM, John Oulton *Foundations of Inference in Natural Science* (Methuen and Co., 1952).
- — "A Reply to Dr. Das's Criticisms," *British Journal for the Philosophy of Science*, Vol. 8 (1957-58), pp. 325-28.
- WORKMAN, Rollin "The Logical Status of the Principle of Induction," *Synthese*, Vol. 13 (1961), pp. 68-74.
- WRIGHT, George H. von. "Carnap's Theory of Probability," *Philosophical Review*, Vol. 60 (1951), pp. 362-74.
- — *A Treatise on Induction and Probability* (New York, Harcourt Brace and Company, 1951).
- — *The Logical Problem of Induction*, Second revised Edition (New York, The Macmillan Company, 1957).
- — "Remarks on the Epistemology of Subjective Probability," Nagel, Suppes and Tarski, *Logic, Methodology and Philosophy of Science*, pp. 330-39.
- ZARAGÜETA, Bengoechea J. "La fonction estimative dans l'induction empirique," *Revue Philosophique de Louvain*, Vol. 49 (1951), pp. 198-201.

## II. "INVESTIGATIONS INTO LOGICAL DEDUCTION"\*

GERHARD GENTZEN

### INTRODUCTION

GERHARD GENTZEN's inaugural dissertation for the University of Göttingen, "*Untersuchungen über das logische Schliessen*," of which Mr. Manfred Szabo here presents an English translation, is an important step in the deeper analysis of the logical calculus as it was begun in the thesis of Jacques Herbrand. In it Gentzen develops a new form of logical calculus, the "calculus of sequents," which he introduces by starting first from a "natural calculus" and by then transforming it for his technical purposes. Its central theorem, an "elimination theorem," is proved in detail in the first part of the paper. The second part deals with applications of this elimination theorem as well as containing the proof of the equivalence of the calculus of sequents with the usual logical calculus.

The *Untersuchungen* have had a profound influence on the development of mathematical logic and proof theory. The treatment of the calculus of sequents has been extended by Oiva Ketonen (in Finland) and by Haskell B. Curry (in the U.S.A.), and the elimination theorem of Gentzen has been transferred by Kurt Schütte to the usual logical calculus. The application of the elimination theorem to proof theory has been developed by Paul Lorenzen and Kurt Schütte through the use of "infinite induction." Lately Gaisi Takeuti (in Japan) has generalized the Gentzen calculus to a system of type theory in which the theory of real numbers can be formalized. Here the elimination theorem is not yet proved, but Takeuti shows that upon the assumption of the generalized elimination theorem the consistency of his system follows.

Thus Gentzen's *Untersuchungen* are of great

current interest, and an English edition of this treatise is to be very much welcomed.

PAUL BERNAYS

(Part I)

### SYNOPSIS

The investigations that follow concern the domain of *predicate logic*, called by H-A<sup>1</sup> the "lower functional calculus." It comprises the types of inference that are continually used in all parts of mathematics. What remains to be added to these are axioms and forms of inference that may be considered as being proper to the particular branches of mathematics, e.g., in elementary number theory the axioms of the natural numbers, of addition, multiplication, and exponentiation, as well as the inference of complete induction; in geometry the geometric axioms.

In addition to *classical logic* I shall also deal with *intuitionist logic* as formalized, for example, by Heyting<sup>2</sup>.

The present investigations into classical and intuitionist predicate logic fall essentially into two only loosely connected parts.

1. My starting point was this: The formalization of logical deduction, especially as it has been developed by Frege, Russell, and Hilbert, is rather far removed from the forms of deduction used in practice in mathematical proofs. Considerable formal advantages are achieved in return. I intended, first of all, to set up a formal system which came as close as possible to actual reasoning. The result was a "*calculus of natural deduction*." ("*NJ*" for intuitionist, "*NK*" for classical predicate logic.) This calculus then turned out to have specific properties; in particular, the "law of the excluded middle," which intuitionists reject, occupies a special position.

\* Originally published in the *Mathematische Zeitschrift*, vol. 39 (1935), pp. 176-221. A second part of these investigations appeared under the same title, *ibid.*, pp. 405-431. The *American Philosophical Quarterly* plans to publish this at a later date. The present English translation is by Mr. M. E. Szabo (McGill University, Montreal) whose thanks are due to his late wife Ann for her continued encouragement, and to Mr. Michael Dummett of All Souls College, Oxford, for reading the translation and suggesting improvements.

<sup>1</sup> Hilbert-Ackermann, *Grundzüge der theoretischen Logik*, in this paper referred to as H-A.

<sup>2</sup> A. Heyting, *Die formalen Regeln der intuitionistischen Logik und Mathematik*, Sitzungsber. d. Preuß. Akad. d. Wiss., phys.-math. Kl. 1930, pp. 42-65.

I shall develop the calculus of natural deduction in the second Section of this paper together with some remarks concerning it.

2. A closer investigation of the specific properties of the natural calculus have finally led me to a very general theorem which will be referred to below as the "*Hauptsatz*."

The *Hauptsatz*<sup>3</sup> says that every purely logical proof can be reduced to a determinate, though not unique, normal form. Perhaps we may express the essential properties of such a normal proof by saying "it is not roundabout." No concepts enter into the proof other than those contained in its final result, and their use was therefore essential to the achievement of that result.

The *Hauptsatz* is valid both for classical and for intuitionist predicate logic.

In order to be able to enunciate and prove the *Hauptsatz* in a convenient form, I had to provide a logical calculus especially suited to the purpose. For this the natural calculus proved unsuitable. For, although it already contains the properties essential to the validity of the *Hauptsatz*, it does so only with respect to its intuitionist form, in view of the fact that the law of excluded middle, as pointed out earlier, occupies a special position in relation to these properties.

In Section III of this paper, therefore, I shall develop a new calculus of logical deduction containing all the desired properties in both their intuitionist and their classical form. ("*LJ*" for intuitionist, "*LK*" for classical predicate logic.) The *Hauptsatz* will then be enunciated and proved by means of that calculus.

The *Hauptsatz* permits of a variety of applications. To illustrate this I shall develop a decision procedure (IV, §1) for intuitionist propositional logic in Section IV, and shall in addition give a new proof of the consistency of classical arithmetic without complete induction (IV, §3).

Sections III and IV may be read independently of Section II.

3. Section I contains the stipulation of the terms and notations used in this paper.

In Section V, I prove the *equivalence* of the logical calculi *NJ*, *NK*, and *LJ*, *LK*, developed in this paper, by means of a calculus modeled on the

formalisms of Russell, Hilbert, and Heyting (and which may easily be compared with them). ("*LHJ*" for intuitionist, "*LHK*" for classical predicate logic.)

4. Only the first part of my paper containing Sections I to III is presented here. Sections IV and V follow in the second part.

## SECTION I.

### STIPULATION OF TERMS

To the concepts "object," "function," "predicate," "proposition," "theorem," "axiom," "proof," "inference," etc., in logic and mathematics there correspond, in the formalization of these disciplines, certain symbols or combinations of symbols. We divide these into:

1. *Symbols*.
2. *Expressions*, i.e., finite series of symbols.
3. *Figures*, i.e., finite sets of symbols, with some ordering.

Symbols count as special cases of expressions and figures, expressions as special cases of figures.

In this paper we shall consider symbols, expressions, and figures of the following kind:

1. *Symbols*.

These divide into constant symbols and variables.

- 1.1 Constant symbols:

*Symbols for particular objects*: 1, 2, 3, . . .

*Symbols for particular functions*: +, −, •.

*Symbols for particular propositions*: **T** ("the true proposition"), **F** ("the false proposition").

*Symbols for particular predicates*: =, <.

*Logical symbols*<sup>4</sup>: & "and," ∨ "or," ⊃ "if . . . then," ⊃ ⊂ "is equivalent to," ⊃ "not", ∀ "for all," ∃ "there is a."

We shall also use the terms: conjunction symbol, disjunction symbol, implication symbol, equivalence symbol, negation symbol, universal quantifier, existential quantifier.

*Auxiliary symbols*: ), (, →.

- 1.2 Variables:

*Object variables*. These we divide into *free* object variables: *a*, *b*, *c*, . . . *m* and *bound* object variables: *n*, . . . *x*, *y*, *z*.

*Propositional variables*: *A*, *B*, *C*, . . .

<sup>3</sup> An important special case of the *Hauptsatz* had already been proved in a completely different way by Herbrand, cf. Section iv, §2.

<sup>4</sup> We take the symbols ∨, ⊃, ∃ from Russell. Russell's symbols for "and," "equivalent," "not," "all," viz: •, ≡, ~, ( ), are already being used with a different meaning in mathematics. We shall therefore take Hilbert's &, whereas Hilbert's symbols for equivalence, all, and not, viz: ~, ( ), ⊃, again have already different meanings. Besides, the negation symbol represents a departure from the linear arrangement of symbols and is inconvenient for some purposes. We shall therefore use Heyting's symbols for equivalence and negation, and for "all" we shall use a symbol [namely ∀] corresponding to ∃.

We assume that an indefinite number of variables is available; if the alphabet does not suffice, we add numerical subscripts such as  $a_7, C_3$ .

1.3. Boldface and Greek letters serve as "syntactic variables," i.e., not as symbols of the logic formalized, but as variables of our considerations concerning that logic. Their meaning is explained as they are used.

## 2. Expressions.

2.1. The concept of a propositional expression, called a "formula" for short (defined inductively):

(The concept of a formula is ordinarily used in a more general sense; the special case defined below might, thus perhaps be described as a "purely logical formula.")

2.1.1. A symbol for a particular proposition (i.e., the symbols **T** and **F**) is a formula.

A propositional variable followed by a number (possibly zero) of free object variables is a formula, e.g.,  $Abab$ .

The object variables are called the *arguments* of the propositional variables.

Formulae of the two kinds mentioned are also called *elementary formulae*.

2.1.2. If **A** is a formula, then  $\neg \mathbf{A}$  is also a formula.

If **A** and **B** are formulae, then  $\mathbf{A} \& \mathbf{B}$ ,  $\mathbf{A} \vee \mathbf{B}$ ,  $\mathbf{A} \supset \mathbf{B}$  are formulae.

(We shall not introduce the symbol  $\supset \subset$  into our presentation; it is in fact superfluous, since  $\mathbf{A} \supset \mathbf{B}$  may be regarded as an abbreviation of  $(\mathbf{A} \supset \mathbf{B}) \& (\mathbf{B} \supset \mathbf{A})$ .)

2.1.3. A formula not containing the bound object variable **x** yields another formula, if we prefix either  $\forall \mathbf{x}$  or  $\exists \mathbf{x}$ . At the same time we may substitute **x** in a number of places for a free object variable occurring in the formula.

2.1.4. Brackets (or parentheses) are to be used to show the structure of a formula unequivocally. Example of a formula:

$$\exists x ((\neg Abxa) \vee Bx) \supset (\forall z (A \& B))$$

By special convention the number of brackets may be reduced, but (with one exception, *vide* 2.4) no use will be made of this, since we do not have to write down many formulae.

2.2. The number of logical symbols occurring in a formula is called the *grade of a formula*. (Thus an elementary formula is of grade 0.)

The logical symbol of a non-elementary formula that has been added last, in the construction of the formula according to 2.1.2 and 2.1.3, is called the *terminal symbol of a formula*.

Formulae that may have arisen in the course of the construction of a formula according to 2.1.2 and 2.1.3, including the formula itself, are called *subformulae*.

Example: the subformulae of  $A \& \forall x Bxa$  are  $A$ ,  $\forall x Bxa$ ,  $A \& \forall x Bxa$  as well as all formulae of the form  $Baa$ , where **a** represents any free object variable (this variable may also be  $a$ , for example). The grade of  $A \& \forall x Bxa$  is 2, the terminal symbol is  $\&$ .

## 2.3. The concept of a sequent:

(This concept will not be used until Section III, and it is only then that the purpose of its introduction becomes apparent.)

A sequent is an expression of the form

$$\mathbf{A}_1, \dots, \mathbf{A}_\mu \longrightarrow \mathbf{B}_1, \dots, \mathbf{B}_\nu$$

where  $\mathbf{A}_1, \dots, \mathbf{A}_\mu, \mathbf{B}_1, \dots, \mathbf{B}_\nu$  may represent any formula whatever. (The  $\longrightarrow$ , like commas, is an auxiliary symbol and not a logical symbol.)

The formulae  $\mathbf{A}_1, \dots, \mathbf{A}_\mu$  forms the *antecedent*, and the formulae  $\mathbf{B}_1, \dots, \mathbf{B}_\nu$  the *succedent* of the sequent. Both expressions may be empty.

2.4. The sequent  $\mathbf{A}_1, \dots, \mathbf{A}_\mu \longrightarrow \mathbf{B}_1, \dots, \mathbf{B}_\nu$  has exactly the same intuitive meaning as the formula

$$(\mathbf{A}_1 \& \dots \& \mathbf{A}_\mu) \supset (\mathbf{B}_1 \vee \dots \vee \mathbf{B}_\nu).$$

(By  $\mathbf{A}_1 \& \mathbf{A}_2 \& \mathbf{A}_3$  we mean  $(\mathbf{A}_1 \& \mathbf{A}_2) \& \mathbf{A}_3$ , likewise for  $\vee$ .)

If the antecedent is empty, the sequent reduces to the formula  $\mathbf{B}_1 \vee \dots \vee \mathbf{B}_\nu$ .

If the succedent is empty, the sequent means the same as the formula  $\neg (\mathbf{A}_1 \& \dots \& \mathbf{A}_\mu)$  or  $(\mathbf{A}_1 \& \dots \& \mathbf{A}_\mu) \supset \mathbf{F}$ .

If both parts of the formula are empty, the sequent means the same as **F**, i.e., a false proposition.

Conversely, to every formula there corresponds an equivalent sequent, e.g., the sequent whose antecedent is empty and whose succedent consists precisely of that formula.

The formulae making up a sequent are called *S-formulae* (i.e., sequent formulae). By this we intend to indicate that we are not considering the formula by itself, but as it appears in the sequent. Thus we say, for example:

"A formula occurs in several places in a sequent as an S-formula"; which may also be expressed as follows:

"Several distinct S-formulae (which shall simply mean: having distinct occurrences in the sequent) are formally identical."

### 3. Figures.

We require inference figures and proof figures.

Such figures consist of formulae or sequents, as the case may be. In what follows (3.1 to 3.3, 3.5) we shall be speaking only of formulae, but whatever is said applies analogously to sequents; all we need to do is to replace the word "formula," wherever it occurs, by the word "sequent."

3.1. An *inference figure* may be written in the following way:

$$\frac{A_1, \dots, A_v}{B} \quad (v \geq 1),$$

where  $A_1, \dots, A_v, B$  are formulae.  $A_1, \dots, A_v$  are then called the *upper formulae* and  $B$  the *lower formula* of the inference figure. (The concepts of an upper sequent and of a lower sequent of an inference figure consisting of sequents are to be understood correspondingly.)

We shall have to consider only particular inference figures and they will be stated for each calculus as they arise.

3.2. A proof figure, called a *derivation* for short, consists of a number of formulae (at least one), which combine to form inference figures in the following way: Each formula is a lower formula of at most one inference figure; each formula (with the exception of exactly one: the *endformula*) is an upper formula of at least one inference figure; and the system of inference figures is non-circular, i.e., there is in the derivation no cycle (no series whose last member is again succeeded by its first member) of formulae of which each upper formula of an inference figure has the lower formula as the next one in the series.

3.3. The formulae of a derivation that are not lower formulae of an inference figure are called *initial formulae* of the derivation.

A derivation is in "tree form" if every one of its formulae is an upper formula of *at most one* inference figure.

Thus all formulae except the endformula are upper formulae of *exactly one* inference figure.

We shall have to treat only of derivations in tree form.

The formulae which compose a derivation so defined are called *D-formulae* (i.e., derivation formulae). By this we wish to indicate that we are not considering merely the formula as such, but also its position in the derivation. In this sense we shall be using, for example, expressions such as:

"A formula occurs in a derivation as a *D-formula*."

"Two distinct *D-formulae* (i.e., formulae occurring merely in distinct places in the derivation) are formally identical, viz., identical to the same formula."

Thus by "*A is the same D-formula as B*" we mean that *A* and *B* are not only formally identical, but occur also in the same place in the derivation. We shall use the words "formally identical" to indicate identity of form regardless of place.

For object variables, however, we shall not introduce a special term that would associate the variable with a specific place of occurrence in the formula. Thus we say, e.g.: "*The same object variable occurs in two distinct D-formulae*."

3.4. The inference figures of the derivation are called *D-inference figures* (i.e., derivation inference figures).

In a derivation consisting of sequents the *S-formulae* of the *D-sequents* are called *D-S-formulae* (i.e., derivation sequent formulae).

3.5. A *branch* in a derivation is (following Hilbert) a series of *D-formulae* whose first formula is an initial formula and whose last formula is the endformula, and of which each formula except the last is an upper formula of a *D-inference figure* whose lower formula is the next formula in the branch.

We say that "a *D-formula* stands *above* (*below*) another *D-formula*" if there exists a branch in which the former occurs before (after) the latter.

We are here thinking of the fact that a derivation is written in tree form with the initial formula above and the endformula below. (Examples may be found in II, §4).

Furthermore, we say that "a *D-inference figure* occurs above (below) a *D-formula*," if all formulae of the inference figure occur above (below) that *D-formula*.

A derivation with the endformula *A* is also called a "derivation of *A*."

The initial formulae of a derivation may be *basic formulae* or *assumption formulae*; more about their nature will have to be said as we reach the different calculi.

## SECTION II.

### THE CALCULUS OF NATURAL DEDUCTION

#### § 1

#### *Examples of Natural Deduction*

We wish to set up a formalism that reflects as accurately as possible the actual logical reasoning involved in mathematical proofs.

By means of a number of examples we shall first of all show what form actual reasoning tends to take and shall examine, for this purpose, three "valid formulae" and try to see their validity in the most natural way possible.

1.1. *First example:*

$(X \vee (Y \& Z)) \supset ((X \vee Y) \& (X \vee Z))$  can be recognized as a valid formula (H-A, p. 28, formula 19).

The argument runs as follows: Suppose that either  $X$  or  $Y \& Z$  holds. We distinguish the two cases: 1.  $X$  holds, 2.  $Y \& Z$  holds. In the first case it follows that  $X \vee Y$  holds, and also  $X \vee Z$ ; hence  $(X \vee Y) \& (X \vee Z)$  also holds. In the second case  $Y \& Z$  holds, which means that both  $Y$  and  $Z$  hold. From  $Y$  follows  $X \vee Y$ ; from  $Z$  follows  $X \vee Z$ . Thus  $(X \vee Y) \& (X \vee Z)$  again holds. The latter formula has thus been derived, in general, from  $X \vee (Y \& Z)$ , i.e.,

$(X \vee (Y \& Z)) \supset ((X \vee Y) \& (X \vee Z))$  holds.

1.2. *Second example:*

$(\exists x \forall y Fxy) \supset (\forall y \exists x Fxy)$ .

(H-A, formula 36, p. 60). The argument runs as follows: Suppose there is an  $x$  such that for all  $y$   $Fxy$  holds. Let  $a$  be such an  $x$ . Then for all  $y$ :  $Fay$  holds. Now let  $b$  be an arbitrary object. Then  $Fab$  holds. Thus there is an  $x$ , viz.,  $a$ , such that  $Fxb$  holds. Since  $b$  was arbitrary, our result therefore holds for all objects, i.e., for all  $y$  there is an  $x$ , such that  $Fxy$  holds. This yields our assertion.

1.3. *Third example:*

$(\neg \exists x Fx) \supset (\forall y \neg Fy)$  is to be recognized as intuitionistically valid. We reason as follows: Assume there were no  $x$  for which  $Fx$  held. From this we wish to infer: For all  $y$ ,  $\neg Fy$  holds. Now suppose  $a$  were some object for which  $Fa$  held. It would then follow that there was an  $x$  for which  $Fx$  held, viz.,  $a$  would be such an object. This contradicts our hypothesis that  $\neg \exists x Fx$ . We have therefore a contradiction, i.e.,  $Fa$  cannot hold. But since  $a$  was completely arbitrary, it follows that for all  $y$ ,  $\neg Fy$  holds. Q.E.D.

We intend now to integrate proofs of the kind carried out in these three examples in an exactly defined calculus (in §4, we shall show how these examples are presented in that calculus).

## §2.

### *Construction of the Calculus NJ*

2.1. We intend now to present a calculus for "natural" intuitionist derivations of valid formulae.

The restriction to intuitionist reasoning is only provisional; we shall explain below (cf. §5) our reasons for doing so and shall show in what way the calculus has to be extended for classical reasoning (by including the law of the excluded middle).

Externally, the essential difference between "NJ-derivations" and derivations in the systems of Russell, Hilbert, and Heyting is the following: In the latter systems true formulae are derived from a series of "logical basic formulae" by means of a few forms of inference. Natural deduction, however, does not, in general, start from logical basic propositions, but rather from *assumptions* (cf. examples in §1) to which logical deductions are applied. By means of a later inference the result is then again made independent of the assumption.

Calculi of the former kind will be referred to as *logistic calculi*.

2.2. After this preliminary remark we define the concept of an *NJ-derivation* as follows:

(Examples in §4).

An *NJ-derivation* consists of formulae ordered in tree form (I, 3.3).

(By demanding that the formulae are in tree form we are deviating somewhat from the analogy with actual reasoning. This is so, since in actual reasoning we necessarily have (1) a linear sequence of propositions due to the linear ordering of our utterances, and (2) we are accustomed to applying repeatedly a result once it has been obtained, whereas the tree form permits only of a single use of a derived formula. These two deviations permit us to define the concept of a derivation in a more convenient form and are not essential.)

The initial formulae of the derivation are *assumption formulae*. Each of these is correlated with precisely one *D-inference* figure (and in fact occurs "above" [I. 3.5] the lower formula of that figure, as will be explained more fully below).

All formulae that occur below an assumption formula, but still above the lower formula of the *D-inference* figure with which that assumption formula is correlated, that assumption formula itself included, are said to be *dependent* on that assumption formula.

(Thus the inference makes all succeeding propositions independent of the assumption which is correlated with it.)

According to what we have said the endformula of the derivation depends on no assumption formula.

2.21. We shall now state the permissible inference figures.

The inference figure schemata below are to be understood in the following way:

We obtain an *NJ*-inference figure from one of the schemata by replacing **A**, **B**, **C**, **D** by arbitrary formulae; and  $\forall \mathbf{x} \mathbf{F} \mathbf{x}$  ( $\exists \mathbf{x} \mathbf{F} \mathbf{x}$ ) by an arbitrary formula containing  $\forall$  or  $\exists$  for its terminal symbol, where  $\mathbf{x}$  indicates the bound object variable belonging to that terminal symbol; and **Fa** by the formula obtained from **Fx** by replacing the bound variable  $\mathbf{x}$ , wherever it occurs, by the free object variable **a**.

(For **a** we may, for instance, take a variable already occurring in **Fx**. For the inference figures  $\forall -I$  and  $\exists -E$ , this possibility will, however, be excluded by the restriction on variables which follows below, but it remains for  $\forall -E$  and  $\exists -I$ . Nor need  $\mathbf{x}$  occur at all in **Fx**, in which case **Fa** is, of course, identical with **Fx**. — **Fa** is obviously always a subformula of  $\forall \mathbf{x} \mathbf{F} \mathbf{x}$  ( $\exists \mathbf{x} \mathbf{F} \mathbf{x}$ ), according to the definition of a subformula in I, 2.2.)

Symbols written in square brackets have the following meaning: An arbitrary number (possibly zero) of formulae of this form, all formally identical, may be correlated with the inference figure as assumption formulae. They must then be initial formulae of the derivation and, moreover, occur in branches of proofs to which the particular upper formula of the inference figure belongs. (I.e., that upper formula above which the square bracket occurs in the scheme. This formula may already itself be an assumption formula.)

The fact that there is a correlation in a derivation between a *D*-inference figure and the related assumption formulae must somehow be made explicit, for example, by jointly numbering them (cf. examples in §4).

The designations of the various inference figure schemata:  $\&-I$ ,  $\&-E$ , etc., stand for the following: An inference figure formed according to a particular schema is an "introduction" (*I*) or an "elimination" (*E*) of the conjunction ( $\&$ ), the disjunction ( $\vee$ ), the universal quantifier ( $\forall$ ), the existential quantifier ( $\exists$ ), the implication ( $\supset$ ), or of the negation ( $\neg$ ). More about this in §5.

The Inference Figure Schemata:

$\&-I$	$\&-E$	$\vee-I$	$\vee-E$
$\frac{\mathbf{A} \quad \mathbf{B}}{\mathbf{A} \& \mathbf{B}}$	$\frac{\mathbf{A} \& \mathbf{B}}{\mathbf{A}} \quad \frac{\mathbf{A} \& \mathbf{B}}{\mathbf{B}}$	$\frac{\mathbf{A} \quad \mathbf{B}}{\mathbf{A} \vee \mathbf{B}} \quad \frac{\mathbf{A} \quad \mathbf{B}}{\mathbf{A} \vee \mathbf{B}}$	$\frac{\mathbf{A} \vee \mathbf{B} \quad \mathbf{C} \quad \mathbf{C}}{\mathbf{C}}$

$\forall -I$	$\forall -E$	$\exists -I$	$\exists -E$
$\frac{\mathbf{F} \mathbf{a}}{\forall \mathbf{x} \mathbf{F} \mathbf{x}}$	$\frac{\forall \mathbf{x} \mathbf{F} \mathbf{x}}{\mathbf{F} \mathbf{a}}$	$\frac{\mathbf{F} \mathbf{a}}{\exists \mathbf{x} \mathbf{F} \mathbf{x}}$	$\frac{\exists \mathbf{x} \mathbf{F} \mathbf{x} \quad \mathbf{C}}{\mathbf{C}}$
$\frac{\supset -I \quad [\mathbf{A}] \quad \mathbf{B}}{\mathbf{A} \supset \mathbf{B}}$	$\frac{\supset -E \quad \mathbf{A} \quad \mathbf{A} \supset \mathbf{B}}{\mathbf{B}}$	$\frac{\neg -I \quad [\mathbf{A}] \quad \mathbf{F}}{\neg \mathbf{A}}$	$\frac{\neg -E \quad \mathbf{A} \quad \neg \mathbf{A}}{\mathbf{F}}$

The free object variable of a  $\forall -I$  or  $\exists -E$ , represented by **a** in the respective schema, is called a *proper variable*. (This, of course, presupposes that there is such a variable, i.e., that the bound object variable represented by  $\mathbf{x}$  occurs in the formula represented by **Fx**.)

*Restrictions on Variables:*

An *NJ*-derivation is subject to the following restriction (for the significance of this restriction cf. §3):

The proper variable of an  $\forall -I$  must not occur in the formula represented in the schema by  $\forall \mathbf{x} \mathbf{F} \mathbf{x}$ , nor in any assumption formula upon which that formula depends.

The proper variable of an  $\exists -E$  must not occur in the formula represented in the schema by  $\exists \mathbf{x} \mathbf{F} \mathbf{x}$ ; nor in an upper formula represented by **C**; nor in any assumption formula upon which that formula depends, with the exception of the assumption formulae represented by **Fa** correlated with the  $\exists -E$ .

This concludes the definition of the "*NJ*-derivation."

### §3.

#### *Intuitive Sense of NJ-Inference Figures*

We shall explain the intuitive sense of a number of inference figure schemata and thus try to show how the calculus in fact reflects "actual reasoning."

$\supset -I$ : Expressed in words, this schema corresponds to the following inference: If **B** has been proved by means of assumption **A**, we have (this time without the assumption): from **A** follows **B**. (Further assumptions may, of course, have been made and the result continues to depend on them.)

$\vee -E$  ("Distinction of Cases"): If  $\mathbf{A} \vee \mathbf{B}$  has been proved, one can distinguish two cases: What one first assumes is that **A** holds and derives, let us say, **C** from it. If it is then possible to derive **C** also by assuming that **B** holds, then **C** holds generally, i.e., it is now independent of both assumptions (cf. 1.1).



inference figure which "introduces" the symbol—as the terminal symbol of a formula—and one which "eliminates" it. The fact that the inference figures  $\&-E$  and  $\vee-I$  each have two forms constitutes a trivial, purely external deviation and is of no interest. The introductions represent, as it were, the "definitions" of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions. This fact may be expressed as follows: In eliminating a symbol, the formula, whose terminal symbol we are dealing with, may be used only "in the sense afforded it by the introduction of that symbol." An example may clarify what is meant: We were able to introduce the formula  $A \supset B$  when there existed a derivation of  $B$  from the assumption formula  $A$ . If we then wished to use that formula by eliminating the  $\supset$ -symbol [we could, of course, also use it to form longer formulae, e.g.,  $(A \supset B) \vee C$ ,  $\vee-I$ ], we could do this precisely by inferring  $B$  directly, once  $A$  has been proved, for what  $A \supset B$  attests is just the existence of a derivation of  $B$  from  $A$ . Note that in saying this we need not go into the "intuitive sense" of the  $\supset$ -symbol.

By making these ideas more precise it should be possible to display the  $E$ -inferences as single-valued functions of their corresponding  $I$ -inferences, on the basis of certain requirements.

5.2. It is possible to eliminate the *negation* from our calculus by regarding  $\neg A$  as an abbreviation of  $A \supset F$ . This is permissible, since by replacing every  $\neg A$  by  $A \supset F$ , and thus removing all  $\neg$ -symbols from an  $\mathcal{NJ}$ -derivation, we obtain another  $\mathcal{NJ}$ -derivation (the inference figures  $\neg-I$  and  $\neg-E$  then become special cases of the  $\supset-I$  and the  $\supset-E$ ) and vice versa: If we replace every occurrence of  $A \supset F$  by  $\neg A$  in an  $\mathcal{NJ}$ -derivation, we obtain another  $\mathcal{NJ}$ -derivation.

#### F

The inference figure schema  $\mathbf{D}$  occupies a special place among the schemata: It does not belong to a logical symbol, but to the propositional symbol  $F$ .

5.3. The "law of the excluded middle" and the calculus  $NK$ .

From the calculus  $\mathcal{NJ}$  we obtain a complete classical calculus  $NK$  by adding the "law of the excluded middle" (*tertium non datur*), i.e.: As initial formulae of a derivation we now also allow in addition to the assumption formulae, "basic formulae" of the form  $A \vee \neg A$ , where  $A$  is to be replaced by an arbitrary formula.

We have thus allotted to the law of the excluded middle, in a purely exterior way, a special position,

and we have done this because we considered that formulation the "most natural." It would be perfectly feasible to introduce a new inference figure schema, say  $\frac{\neg \neg A}{A}$  (a schema analogous to the one formed by Hilbert and Heyting) in place of the basic formula schema  $A \vee \neg A$ . However, such a schema still falls outside the framework of the  $\mathcal{NJ}$ -inference figures, because it represents a new elimination of the negation whose admissibility does not follow at all from our method of introducing the  $\neg$ -symbol by the  $\neg-I$ .

### SECTION III.

#### DEDUCTIVE CALCULI LJ, LK AND THE HAUPTSATZ

##### §1.

##### *The Calculi LJ and LK*

##### *(Logistic Intuitionist and Classical Calculus)*

1.1. Preliminary remarks concerning the construction of the calculi  $LJ$  and  $LK$ .

What we want to do is to formulate a deductive calculus (for predicate logic) which is "logistic" on the one hand, i.e., in which the derivations do not, as in the calculus  $\mathcal{NJ}$ , contain assumption formulae, but which, on the other hand, takes over from the calculus  $\mathcal{NJ}$  the division of the forms of inference into introductions and eliminations of the various logical symbols.

The most obvious method of converting an  $\mathcal{NJ}$ -derivation into a logistic one is this: We replace a  $D$ -formula  $B$ , which depends on the assumption formulae  $A_1, \dots, A_n$ , by the new formula  $(A_1 \& \dots \& A_n) \supset B$ . This we do in all  $D$ -formulae.

We thus obtain formulae which are already valid *in themselves*, i.e., whose truth is no longer *conditional* on the truth of certain assumption formulae. This procedure, however, introduces the new logical symbols  $\&$  and  $\supset$ , necessitating additional inference figures for  $\&$  and  $\supset$ , and thus upsets the systematic character of our method of introducing and eliminating symbols. For this reason we have introduced the concept of a *sequent* (I, 2.3). Instead of the formula  $(A_1 \& \dots \& A_n) \supset B$ , e.g., we therefore write the sequent

$$A_1, \dots, A_n \multimap B$$

The sequent does not distinguish itself from the above formula in its intuitive meaning, but only in its formal structure (cf. I, 2.4).

Even now new inference figures are required that cannot be integrated into our system of introductions and eliminations; but we have the advantage of being able to reserve them special places within our system, since they no longer refer to logical symbols, but merely to the structure of the sequents. We therefore call these "structural inference figures," and the others "operational inference figures."

In the classical calculus *NK* the law of the excluded middle occupied a special place among the forms of inference (II, 5.3), because it could not be integrated into our system of introductions and eliminations. In the classical logistic calculus *LK* to be presented below, that peculiarity is removed. What makes this possible is that we admit into our system sequents with several formulae in the succedent, whereas the indicated transition from the calculus *NJ* has resulted only in sequents with one formula in the succedent. (For the intuitive meaning of the general sequents cf. I, 2.4.) The symmetry thus obtained is more suited to classical logic. On the other hand, the restriction to at most one formula in the succedent will be retained for the intuitionist calculus *LJ*. (Cf. below —An empty succedent means the same as if *F* stood in the succedent.)

We have thus outlined a number of points that underlie the construction of the calculi that follow. Their form is largely determined, however, by considerations connected with the "*Hauptsatz*" (§2) whose proof follows later. That form cannot therefore be justified more fully at this stage.

1.2. We now define the concepts of a "*LK*-derivation" and a "*LJ*-derivation" as follows:

An *LJ*- or *LK*-derivation consists of sequents arranged in tree form (I, 3.3).

The initial sequents of the derivation are basic sequents of the form

$$D \rightarrow D,$$

where *D* may be an arbitrary formula.

Each inference figure of the derivation results from one of the schemata below by a substitution of the following kind (cf. II, 2.21):

Replace *A*, *B*, *D*, *E* by an arbitrary formula; for  $\forall xFx$  ( $\exists xFx$ ) put an arbitrary formula having  $\forall$  ( $\exists$ ) for its terminal symbol, where *x* designates the associated bound object variable; for *Fa* put that formula which is obtained from *Fx* by replacing every occurrence of the bound object variable *x* by the free object variable *a*.

For  $\Gamma$ ,  $\Delta$ ,  $\Theta$ ,  $\Lambda$  put arbitrary (possibly empty) sequences of formulae separated by commas.

The following restriction is furthermore placed on *LJ*-inference figures (this is the only respect in which the concepts of a *LJ*- and a *LK*-derivation differ):

"In the succedent of each *D*-sequent no more than one *S*-formula is permissible."

The designations of the various schemata for operational inference figures &-IS, &-IA, etc., are intended to mean: An inference figure formed according to the schema is an introduction (*I*) in the succedent (*S*) or antecedent (*A*) of the conjunction (&), the disjunction ( $\vee$ ), the universal quantifier ( $\forall$ ), the existential quantifier ( $\exists$ ), the negation ( $\neg$ ), or the implication ( $\rightarrow$ ).

### The Inference Figure Schemata

1.21. Schemata for structural inference figures: Thinning:

$$\begin{array}{cc} \text{in the antecedent} & \text{in the succedent} \\ \frac{\Gamma \rightarrow \Theta}{D, \Gamma \rightarrow \Theta'} & \frac{\Gamma \rightarrow \Theta}{\Gamma \rightarrow \Theta, D'} \end{array}$$

Contraction:

$$\begin{array}{cc} \text{in the antecedent} & \text{in the succedent} \\ \frac{D, D, \Gamma \rightarrow \Theta}{D, \Gamma \rightarrow \Theta'} & \frac{\Gamma \rightarrow \Theta, D, D}{\Gamma \rightarrow \Theta, D} \end{array}$$

Interchange:

$$\begin{array}{cc} \text{in the antecedent} & \text{in the succedent} \\ \frac{\Delta, D, E, \Gamma \rightarrow \Theta}{\Delta, E, D, \Gamma \rightarrow \Theta'} & \frac{\Gamma \rightarrow \Theta, E, D, \Lambda}{\Gamma \rightarrow \Theta, D, E, \Lambda} \end{array}$$

Cut:

$$\frac{\Gamma \rightarrow \Theta, D \quad D, \Delta \rightarrow \Lambda}{\Gamma, \Delta \rightarrow \Theta, \Lambda}$$

1.22. Schemata for operational inference figures:

$$\&-IS: \frac{\Gamma \rightarrow \Theta, A \quad \Gamma \rightarrow \Theta, B}{\Gamma \rightarrow \Theta, A \& B}$$

$$\&-IA: \frac{A, \Gamma \rightarrow \Theta \quad B, \Gamma \rightarrow \Theta}{A \& B, \Gamma \rightarrow \Theta}$$

$$\vee-IS: \frac{A, \Gamma \rightarrow \Theta \quad B, \Gamma \rightarrow \Theta}{A \vee B, \Gamma \rightarrow \Theta}$$

$$\vee-IA: \frac{\Gamma \rightarrow \Theta, A \quad \Gamma \rightarrow \Theta, B}{\Gamma \rightarrow \Theta, A \vee B}$$

$$\forall\text{-IS: } \frac{\Gamma \rightarrow \theta, Fa}{\Gamma \rightarrow \theta, \forall xFx}$$

$$\exists\text{-IA: } \frac{Fa, \Gamma \rightarrow \theta}{\exists xFx, \Gamma \rightarrow \theta}$$

*Restrictions on Variables:* The object variable in the last two schemata, which is designated by **a** and is called the *proper variable* of the  $\forall\text{-IS}$  ( $\exists\text{-IA}$ ), must not occur in the lower sequent of the inference figure (i.e., not in  $\Gamma$ ,  $\theta$ , and  $Fx$ ).

$$\forall\text{-IA: } \frac{Fa, \Gamma \rightarrow \theta}{\forall xFx, \Gamma \rightarrow \theta}$$

$$\neg\text{-IS: } \frac{A, \Gamma \rightarrow \theta}{\Gamma \rightarrow \theta, \neg A}$$

$$\exists\text{-IS: } \frac{\Gamma \rightarrow \theta, Fa}{\Gamma \rightarrow \theta, \exists xFx}$$

$$\neg\text{-IA: } \frac{\Gamma \rightarrow \theta, A}{\neg A, \Gamma \rightarrow \theta}$$

$$\supset\text{-IS: } \frac{A, \Gamma \rightarrow \theta, B}{\Gamma \rightarrow \theta, A \supset B}$$

$$\supset\text{-IA: } \frac{\Gamma \rightarrow \theta, A \quad B, \Delta \rightarrow A}{A \supset B, \Gamma, \Delta \rightarrow \theta, A}$$

1.3. Example of an  $LJ$ -derivation (following II, 4.3):

$$\begin{array}{c} \frac{Fa \rightarrow Fa}{Fa \rightarrow \exists xFx} \quad \exists\text{-IS} \quad \frac{\exists xFx \rightarrow \exists xFx}{\neg \exists xFx, \exists xFx \rightarrow} \quad \neg\text{-IA} \\ \frac{Fa, \neg \exists xFx \rightarrow}{\neg \exists xFx \rightarrow \neg Fa} \quad \text{Interchange} \\ \frac{\neg \exists xFx \rightarrow \neg Fa}{\neg \exists xFx \rightarrow \forall y \neg Fy} \quad \neg\text{-IS} \\ \frac{\neg \exists xFx \rightarrow \forall y \neg Fy}{\rightarrow (\neg \exists xFx) \supset (\forall y \neg Fy)} \quad \supset\text{-IS} \end{array}$$

1.4. Example of an  $LK$ -derivation (derivation of the "law of the excluded middle"):

$$\begin{array}{c} \frac{A \rightarrow A}{\rightarrow A, \neg A} \quad \neg\text{-IS} \\ \frac{\rightarrow A, \neg A}{\rightarrow A, A \vee \neg A} \quad \vee\text{-IS} \\ \frac{\rightarrow A, A \vee \neg A}{\rightarrow A \vee \neg A, A} \quad \text{Interchange} \\ \frac{\rightarrow A \vee \neg A, A \vee \neg A}{\rightarrow A \vee \neg A} \quad \vee\text{-IS} \\ \frac{\rightarrow A \vee \neg A}{\rightarrow A \vee \neg A} \quad \text{Contraction} \end{array}$$

D

§2.

### Some Remarks Concerning the Calculi $LJ$ and $LK$

#### The Hauptsatz

(We shall make no further use, in this paper, of remarks 2.1 to 2.3.)

2.1. The schemata are not all mutually independent, i.e., certain schemata could be eliminated with the help of the remaining ones. Yet if they were left out, the "*Hauptsatz*" would no longer be valid.

2.2. In general, we could *simplify* the calculi in various respects if we attached no importance to the *Hauptsatz*. To indicate this briefly: the inference figures  $\&\text{-IS}$ ,  $\vee\text{-IA}$ ,  $\&\text{-IA}$ ,  $\vee\text{-IS}$ ,  $\forall\text{-IA}$ ,  $\exists\text{-IS}$ ,  $\neg\text{-IS}$ ,  $\neg\text{-IA}$ , and  $\supset\text{-IA}$  in the calculus  $LK$  could be replaced by basic sequents according to the following schemata:

$$\begin{array}{l} A, B \rightarrow A \& B \quad A \vee B \rightarrow A, B \quad A \& B \rightarrow A \\ A \& B \rightarrow B \quad A \rightarrow A \vee B \quad B \rightarrow A \vee B \quad \forall xFx \rightarrow Fa \\ Fa \rightarrow \exists xFx \quad \rightarrow A, \neg A \quad (\text{law of the excluded middle}) \end{array}$$

$$\neg A, A \rightarrow (\text{law of contradiction}), A \supset B, A \rightarrow B$$

These basic sequents and our inference figures may easily be shown to be equivalent.

The same possibility exists for the calculus  $LJ$ , with the exception of the inference figures  $\vee\text{-IA}$  and  $\neg\text{-IS}$ , since  $LJ$ - $D$ -sequents may not in fact contain two  $S$ -formulae in the succedent (cf. V, §5).

2.3. The distinction between *intuitionist* and *classical* logic is, externally, of a quite different nature in the calculi  $LJ$  and  $LK$  from that in the calculi  $NJ$  and  $NK$ . In the case of the latter, the distinction rests on the inclusion or exclusion of the law of the excluded middle, whereas for the calculi  $LJ$  and  $LK$  the difference is characterized by the restrictions on the succedent. (The fact that both distinctions are equivalent will become evident as a result of the equivalence proofs in Section V for all calculi discussed in this paper.)

2.4. If  $\supset\text{-IS}$  and the  $\supset\text{-IA}$  are excluded, the calculus  $LK$  is *dual* in the following sense: If we reverse all sequents of an  $LK$ -derivation (in which the  $\supset$ -symbol does not occur), i.e., if for  $A_1, \dots, A_n \rightarrow B_1, \dots, B_m$  we put  $B_1, \dots, B_m \rightarrow A_1, \dots, A_n$ ; and if we exchange, in inference figures with two upper sequents, the right and left-hand upper sequents, including their derivations, and also replace every occurrence of  $\&$  by  $\vee$ ,  $\forall$  by  $\exists$ ,  $\vee$  by  $\&$ , and  $\exists$  by  $\forall$  (in the case of  $\&$  and  $\vee$  we also

have to change the respective scope of the symbols, e.g., for  $B \vee A$  we have to put  $A \& B$ , then another  $LK$ -derivation results.

This can be seen at once from the schemata. (Special care was taken to order them in such a way as to bring out their symmetry.)

(Cf. H-A's duality principle, p. 62).

2.41. In any case, the  $\supset$ -symbol may, in a well-known manner, be eliminated from the calculus  $NK$ , by regarding  $A \supset B$  as an abbreviation for  $(\neg A) \vee B$ . It may easily be shown that the schemata for the  $\supset$ -IS and the  $\supset$ -IA may then be replaced by the schemata for  $\vee$  and  $\neg$ .

The calculus  $NJ$  has no corresponding property.

2.5. The most important fact for us with regard to the calculi  $LJ$  and  $LK$  is the following:

HAUPTSATZ: Every  $LJ$ - or  $LK$ -derivation can be transformed into an  $LJ$ - or  $LK$ -derivation that has the same endsequent and in which the inference figure termed a "cut" does not occur.

2.51. The proof follows in §3.

In order to give greater clarity to the meaning of the *Hauptsatz*, we prove a simple *corollary* (2.513).

For this purpose we introduce a number of expressions (which will be needed frequently later on) relating to operational inference figures.

2.511. That  $S$ -formula which contains the logical symbol in its schema will be called the *principal formula* of an inference figure.

For the  $\&$ -IS and the  $\&$ -IA this is simply the  $S$ -formula of the form  $A \& B$ ; for the  $\vee$ -IS and the  $\vee$ -IA it is  $A \vee B$ ; for the  $\forall$ -IS and the  $\forall$ -IA it is  $\forall x Fx$ ; for the  $\exists$ -IS and the  $\exists$ -IA it is  $\exists x Fx$ ; for the  $\neg$ -IS and the  $\neg$ -IA it is  $\neg A$ ; and for the  $\supset$ -IS and the  $\supset$ -IA it is  $A \supset B$ .

The  $S$ -formulae designated by  $A$ ,  $B$ ,  $Fa$  in the schemata we call *side formulae* of their corresponding inference figures.

They are always subformulae of the principal formula (according to the definition of a subformula in I, 2.2).

2.512. We can now easily read off the following facts from the inference figure schemata: A principal formula occurs always in the lower sequent, a side formula always in the upper sequents of an operational inference figure.

In the case of a formula occurring as an  $S$ -formula in an upper sequent of an arbitrary inference figure without being a side formula or the  $D$  of a cut, it occurs also in the lower sequent as an  $S$ -formula.

These two facts entail the following:

If anywhere in an  $LJ$ - or  $LK$ -derivation a

formula occurs as an  $S$ -formula, and if we trace the branch of the derivation from the formula concerned up to the endsequent, the formula can only then vanish from that branch if it is the  $D$  of a cut or the side formula of an operational inference figure. In the latter case, however, there appears, in the next sequent, the *principal formula* of the inference figure of which our side formula is a subformula. To that principal formula we can then, continuing downwards, apply the same consideration, and so on. Thus we obtain the following corollary:

2.513. *Corollary of the Hauptsatz (subformula property)*: In an  $LJ$ - or  $LK$ -derivation without cuts, all occurring  $D$ - $S$ -formulae are *subformulae* of the  $S$ -formula that occurs in the endsequent.

2.514. Intuitively speaking, these properties of derivations without cuts may be expressed as follows: The  $S$ -formulae become larger as we descend lower down in the derivation, never shorter. The final result is, as it were, gradually built up from its constituent elements. The proof represented by the derivation is not roundabout in that it contains only concepts which recur in the final result (cf. the synopsis at the beginning of this paper).

*Example*: The derivation given above (1.3) for  $\rightarrow (\neg \exists x Fx) \supset (\forall y \neg Fy)$  may be written without a cut as follows:

$$\frac{\frac{Fa \rightarrow Fa}{Fa \rightarrow \exists x Fx} \exists\text{-IS}}{\neg \exists x Fx, Fa \rightarrow} \neg\text{-IA} \\ \frac{}{Fa, \neg \exists x Fx \rightarrow} \text{Interchange}$$

etc., as above.

### §3.

#### *Proof of the Hauptsatz*

The *Hauptsatz* runs as follows:

Every  $LJ$ - or  $LK$ -derivation can be transformed into another  $LJ$ - or  $LK$ -derivation with the same endsequent, in which no cuts occur.

#### 3.1. *Proof of the Hauptsatz for $LK$ -derivations.*

We introduce a new inference figure (in order to facilitate the proof) that constitutes a modified form of the cut, and which we call a *mix*.

The schema of that figure runs as follows:

$$\frac{\Gamma \rightarrow \Theta \quad \Delta \rightarrow A}{\Gamma, \Delta^* \rightarrow \Theta^*, A}$$

In order to obtain an inference figure from this schema,  $\Theta$  and  $\Delta$  must be replaced by sequences of

formulae, separated by commas, in each of which occurs at least once (as a member of the sequence) a formula of the form  $\mathbf{M}$ , called the "mix formula"; and  $\Theta^*$  and  $\Delta^*$  must be replaced by the same sequences of formulae, save that all formulae of the form  $\mathbf{M}$  occurring as members of the sequence are omitted. ( $\mathbf{M}$  is an arbitrary formula.)  $\Gamma$  and  $\Delta$  must be replaced, as in the other schemata, by arbitrary (possibly empty) sequences of formulae, separated by commas.

Example of a mix:

$$\frac{A \rightarrow B, \neg A \quad B \vee C, B, B, D, B \rightarrow}{A, B \vee C, D \rightarrow \neg A}$$

$B$  is the mix formula.

We notice at once that every cut may be transformed into a mix by means of a number of thinnings and interchanges. (Conversely, every mix may be transformed into a cut by means of a certain number of preceding interchanges and contractions, though we do not use this fact.)

In the following we shall consider only derivations in which no cuts occur, but which may contain mixes instead.

Since derivations in the old sense may be transformed into derivations of the new kind, it suffices, for the proof of the *Hauptsatz*, to show that a derivation of the new type may be transformed into a derivation with no mix.

Furthermore, the following lemma is already sufficient:

*Lemma:* Any derivation with a mix for its lowest inference figure, and not containing any other mix, may be transformed into a derivation (with the same endsequent) in which no mix occurs.

From this the complete theorem easily follows:

In an arbitrary derivation consider a mix above whose lower sequent no further mix occurs. The derivation for this lower sequent is then of the kind mentioned in the lemma, i.e., it may be transformed in such a way that it no longer contains a mix. In doing so, the rest of the derivation remains unchanged. This operation is then repeated until every mix has systematically been eliminated.

It now remains for us to establish the *proof of the lemma*. (This proof extends into 3.2 incl.)

We have to consider a derivation whose lowest inference figure is a mix and which contains no other mix besides.

The grade of the mix formula will be called the "grade of the derivation" (defined in I, 2.2).

We shall call "rank of the derivation" the sum of its rank on the left and its rank on the right. These two terms are defined as follows:

The left rank is the greatest number of sequents connected on a branch such that the last sequent is the *left-hand* upper sequent of the mix, where each formula of the branch contains the mix formula in the *succedent*.

The right rank is (correspondingly) the greatest number of sequents connected on a branch whose lowest sequent is the *right-hand* upper sequent of the mix, and each formula of the branch contains the mix formula in the *antecedent*.

The lowest possible rank is evidently 2.

To prove the lemma we perform two complete inductions, one according to the grade  $\gamma$ , the other according to the rank  $\rho$  of the derivation, i.e., we prove the theorem for a derivation of the grade  $\gamma$ , assuming it to hold for derivations of a lower grade (in so far as there are such derivations, i.e., as long as  $\gamma$  is not equal to zero), supposing, therefore, that derivations of a lower grade may already be transformed into derivations not containing a mix.

Furthermore, we shall begin by considering the case where the rank  $\rho$  of the derivation equals 2 (3.11), and after that the case of  $\rho > 2$  (3.12), where we assume that the theorem already holds for derivations of the same grade, but of a lower rank.

In the following bold-face capital letters will generally serve as syntactic variables for *formulae*, and Greek capital letters as syntactic variables for (possibly empty) *sequences of formulae*.

In transforming derivations, we shall occasionally meet "identical inference figures," i.e., inference figures with the same upper and lower sequent. Since we have not admitted such figures in our calculus, they must be eliminated as soon as they occur; this is trivially possible by omitting one of the two sequents.

The mix formula of the mix that occurs at the end of the derivation is designated by  $\mathbf{M}$ . It is of grade  $\gamma$ .

3.10. *Re-designating of free object variables* in preparation for the transformation of derivations.

We wish to obtain a derivation that has the following properties:

3.101. For every  $\forall\text{-IS}$  ( $\exists\text{-IA}$ ) it holds that: Its proper variable occurs in the derivation only in sequents *above* the lower sequent of the  $\forall\text{-IS}$  ( $\exists\text{-IA}$ ) and does not occur as a proper variable in any other  $\forall\text{-IS}$  ( $\exists\text{-IA}$ ).

3.102. This is achieved by re-designating the free object variables in the following way:

We take a  $\forall$ -IS ( $\exists$ -IA) above whose lower sequent either no further inference figures of this kind occur, or if they do, they have already been dealt with in a way to be outlined.

In all sequents above the lower sequent of this inference figure we replace the proper variables by one and the same free object variable which, so far, has not yet occurred in the derivation. This obviously leaves the validity of the  $\forall$ -IS ( $\exists$ -IA) as such unchanged. (The proper variables did in fact not occur in its lower sequent.) Furthermore, the remainder of the derivation remains correct, as is shown by the immediately following lemma:

By applying this method systematically to every single  $\forall$ -IS and  $\exists$ -IA, the derivation thus remains correct throughout and at the conclusion obviously has the desired property (3.101). Furthermore, as was essential, the grade and rank of the derivation as well as its endsequent have remained unaltered.

3.103. Now we give the still outstanding proof of the following lemma. (It is enunciated in a somewhat more general form than is immediately necessary, since we shall have to apply it again later on (3.113.33).)

"An LK-basic sequent or inference figure turns into a basic sequent or inference figure of the same kind, if we replace a free object variable, which is *not* the proper variable of the inference figure, in all its occurrences in the basic sequent or inference figure, by one and the same free object variable, provided again that that is *not* the proper variable of the inference figure."

This holds trivially except for the  $\forall$ -IS, the  $\forall$ -IA, the  $\exists$ -IS and the  $\exists$ -IA. Yet even here there is no cause for concern: the restrictions on variables are not violated, since we may neither substitute nor replace the proper variable. (This is the reason why both restrictions on variables are necessary.) Furthermore, the formula resulting from  $F\mathbf{a}$  is still obtained by substituting  $\mathbf{a}$  for  $\mathbf{x}$  in the formula resulting from  $F\mathbf{x}$ .

Having prepared the way (3.10), we now proceed to the actual transformation of the derivation which serves to eliminate the mix occurring in it.

As already mentioned, we distinguish two cases:  $\rho = 2$  (3.11) and  $\rho > 2$  (3.12).

3.11. Suppose  $\rho = 2$ .

We distinguish between a number of particular cases, of which cases 3.111, 3.112, 3.113.1, 3.113.2

are especially simple in that they allow the mix to be immediately eliminated. The other cases (3.113.3) are the most important since their consideration brings out the basic idea behind the whole transformation. Here we use the induction hypothesis with respect to  $\gamma$ , i.e., we reduce each one of the cases to transformed derivations of a lower grade.

3.111. Suppose the left-hand upper sequent of the mix at the end of the derivation is a *basic sequent*. The mix then reads:

$$\frac{\mathbf{M} \rightarrow \mathbf{M} \quad \Delta \rightarrow \mathbf{A}}{\mathbf{M}, \Delta^* \rightarrow \mathbf{A}}$$

which is transformed into:

$$\frac{\Delta \rightarrow \mathbf{A}}{\mathbf{M}, \Delta^* \rightarrow \mathbf{A}} \quad \text{possibly several interchanges and contractions.}$$

That part of the derivation which is above  $\Delta \rightarrow \mathbf{A}$  remains the same, and we thus already have a derivation without a mix.

3.112. Suppose the right-hand upper sequent of the mix is a *basic sequent*. The treatment of this case is symmetric to that of the previous one. We have only to regard the two schemata as duals (cf. 2.4).

3.113. Suppose that neither the left- nor the right-hand upper sequent of the mix is a basic sequent. Then both are *lower sequents of inference figures* since  $\rho = 2$ , and the right and left rank both equal 1, i.e.: In the sequents directly above the *left-hand* upper sequent of the mix, the mix formula  $\mathbf{M}$  does not occur in the *succedent*; in the sequents directly above the *right-hand* upper sequent  $\mathbf{M}$  does not occur in the *antecedent*.

Now the following holds generally: If a formula occurs in the antecedent (succedent) of the lower sequent of an inference figure, it is either a principal formula or the  $\mathbf{D}$  of a thinning, or else it also occurs in the antecedent (succedent) in at least one upper sequent of the inference figure.

This can be seen immediately by looking at the inference figure schemata (1.21, 1.22).

If we now consider the assumptions of the following three cases, we see at once that they exhaust all the possibilities that exist within case 3.113.

3.113.1. Suppose the left-hand upper sequent of the mix is the lower sequent of a *thinning*. Then the conclusion of the derivation runs:

$$\frac{\Gamma \rightarrow \Theta}{\Gamma \rightarrow \Theta, \mathbf{M} \quad \Delta \rightarrow \mathbf{A}} \quad \frac{\Gamma \rightarrow \Theta, \mathbf{M} \quad \Delta \rightarrow \mathbf{A}}{\Gamma, \Delta^* \rightarrow \Theta, \mathbf{A}}$$

This is transformed into:

$$\frac{\Gamma \rightarrow \Theta}{\Gamma, \Delta^* \rightarrow \Theta, \Delta} \text{ possibly several thinnings and interchanges.}$$

That part of the derivation which occurs above  $\Delta \rightarrow \Delta$  disappears.

3.113.2. Suppose the right-hand upper sequent of the mix is the lower sequent of a thinning. This case can be dealt with symmetrically to the previous one.

3.113.3. The mix formula **M** occurs both in the succedent of the left-hand upper sequent and in the antecedent of the right-hand upper sequent solely as the *principal formula* of one of the operational inference figures.

Depending on whether the terminal symbol of **M** is  $\&$ ,  $\vee$ ,  $\forall$ ,  $\exists$ ,  $\neg$ ,  $\supset$ , we distinguish the cases 3.113.31 to 3.113.36 (a formula without logical symbols cannot be a principal formula).

3.113.31. Suppose the terminal symbol of **M** is  $\&$ . In that case the end of the derivation runs:

$$\frac{\Gamma_1 \rightarrow \Theta_1, \mathbf{A} \quad \Gamma_1 \rightarrow \Theta_1, \mathbf{B} \quad \&-IS \quad \mathbf{A}, \Gamma_2 \rightarrow \Theta_2 \quad \&-IA}{\Gamma_1 \rightarrow \Theta_1, \mathbf{A} \& \mathbf{B} \quad \mathbf{A} \& \mathbf{B}, \Gamma_2 \rightarrow \Theta_2 \text{ mix}} \Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2$$

(and correspondingly for the other form of the  $\&-IA$ , treated analogously.)

We transform it into:

$$\frac{\Gamma_1 \rightarrow \Theta_1, \mathbf{A} \quad \mathbf{A}, \Gamma_2 \rightarrow \Theta_2 \text{ mix}}{\Gamma_1, \Gamma_2^* \rightarrow \Theta_1^*, \Theta_2 \text{ possibly several thinnings}} \Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2 \text{ and interchanges.}$$

We can now apply the induction hypothesis with respect to  $\gamma$  to that part of the derivation whose lowest sequent is  $\Gamma_1, \Gamma_2^* \rightarrow \Theta_1^*, \Theta_2$ , because it has a lower grade than  $\gamma$ . (**A** obviously contains fewer logical symbols than **A** & **B**.) This means that the whole derivation may be transformed into one with no mix.

3.113.32. Suppose the terminal symbol of **M** is  $\vee$ . This case is to be dealt with symmetrically to the previous one.

3.113.33. Suppose the terminal symbol of **M** is  $\forall$ . Then the end of the derivation runs:

$$\frac{\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{a} \quad \forall-IS \quad \mathbf{F}\mathbf{b}, \Gamma_2 \rightarrow \Theta_2 \quad \forall-IA}{\Gamma_1 \rightarrow \Theta_1, \forall \mathbf{x} \mathbf{F}\mathbf{x} \quad \forall \mathbf{x} \mathbf{F}\mathbf{x}, \Gamma_2 \rightarrow \Theta_2 \text{ mix}} \Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2$$

This is transformed into:

$$\frac{\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{b} \quad \mathbf{F}\mathbf{b}, \Gamma_2 \rightarrow \Theta_2 \text{ mix}}{\Gamma_1, \Gamma_2^* \rightarrow \Theta_1^*, \Theta_2 \text{ possibly several thinnings}} \Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2 \text{ and interchanges.}$$

Above the left-hand upper sequent of the mix,  $\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{b}$ , we write the same part of the derivation which previously occurred above  $\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{a}$ , yet having replaced every occurrence of the free object variable **a** by **b**. It now follows from the lemma 3.103, together with 3.101, that in performing this operation the part of the derivation above  $\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{b}$  has again become a correct part of the derivation. (By virtue of 3.101 neither **a** nor **b** can be the proper variable of an inference figure occurring in that part of the derivation.) The same consideration may be applied to that part of the derivation which includes the sequent  $\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{b}$ , since it too results from  $\Gamma_1 \rightarrow \Theta_1, \mathbf{F}\mathbf{a}$  by substitution of **b** for **a**. It is now in fact clear that by virtue of the restriction on variables for  $\forall-IS$ , **a** could have occurred neither in  $\Gamma_1$  and  $\Theta_1$ , nor in **Fx**. Furthermore, **Fa** results from **Fx** by substituting **a** for **x**, and **Fb** from **Fx** by substituting **b** for **x**. This is why **Fb** results from **Fa** by substituting **b** for **a**.

The mix formula **Fb** in the new derivation has a lower grade than  $\gamma$ . Therefore, according to the induction hypothesis, the mix may be eliminated.

3.113.34. Suppose the terminal symbol of **M** is  $\exists$ . This case is resolved symmetrically to the previous one.

3.113.35. Suppose the terminal symbol of **M** is  $\neg$ . Then the end of the derivation runs:

$$\frac{\mathbf{A}, \Gamma_1 \rightarrow \Theta_1 \quad \neg-IS \quad \Gamma_2 \rightarrow \Theta_2, \mathbf{A} \quad \neg-IA}{\Gamma_1 \rightarrow \Theta_1, \neg \mathbf{A} \quad \neg \mathbf{A}, \Gamma_2 \rightarrow \Theta_2 \text{ mix}} \Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2$$

This is transformed into:

$$\frac{\Gamma_2 \rightarrow \Theta_2, \mathbf{A} \quad \mathbf{A}, \Gamma_1 \rightarrow \Theta_1 \text{ mix}}{\Gamma_2, \Gamma_1^* \rightarrow \Theta_2^*, \Theta_1 \text{ possibly several interchanges and thinnings}} \Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2$$

The new mix may be eliminated by virtue of the induction hypothesis.

3.113.36. Suppose the terminal symbol of **M** is  $\supset$ . Then the end of the derivation runs:

$$\frac{\frac{A, \Gamma_1 \rightarrow \theta_1, B}{\Gamma_1 \rightarrow \theta_1, A \supset B} \supset -IS \quad \frac{\Gamma \rightarrow \theta, A \quad B, \Delta \rightarrow A}{A \supset B, \Gamma, \Delta \rightarrow \theta, A} \supset -IA}{\Gamma_1, \Gamma, \Delta \rightarrow \theta_1, \theta, A} \text{ mix}$$

This is transformed into:

$$\frac{\frac{A, \Gamma_1 \rightarrow \theta_1, B \quad B, \Delta \rightarrow A}{\Gamma \rightarrow \theta, A \quad A, \Gamma_1, \Delta^* \rightarrow \theta_1^*, A} \text{ mix}}{\frac{\Gamma, \Gamma_1^*, \Delta^{**} \rightarrow \theta^*, \theta_1^*, A}{\Gamma_1, \Gamma, \Delta \rightarrow \theta_1, \theta, A} \text{ possibly several interchanges and thinnings}} \text{ mix}$$

(The asterisks are, of course, intended to be understood as follows:  $\Delta^*$  and  $\theta^*$  result from  $\Delta$  and  $\theta_1$  by omitting all  $S$ -formulae of the form  $B$ ;  $\Gamma_1^*$ ,  $\Delta^{**}$  and  $\theta^*$  result from  $\Gamma_1$ ,  $\Delta^*$  and  $\theta$  by omitting all  $S$ -formulae of the form  $A$ .)

Now we have two mixes, but both mix formulae are of a lower grade than  $\gamma$ . We first apply the induction hypothesis to the upper mix (i.e., to that part of the derivation whose lowest figure it is). Thus the upper mix may be eliminated. We can then also eliminate the lower mix.

3.12. Suppose  $\rho > 2$ .

To begin with, we distinguish two main cases: First case: The right rank is greater than 1 (3.121). Second case: The right rank is equal to 1 and the left rank is therefore greater than 1 (3.122). The second case may essentially be dealt with symmetrically to the first.

3.121. Suppose the right rank is greater than 1.

I.e.: The right-hand upper sequent of the mix is the lower sequent of an inference figure, let us call it **If**, and **M** occurs in the antecedent of at least one upper sequent of **If**.

The basic idea behind the transformation procedure is the following:

In the case of  $\rho = 2$ , we generally reduced the derivation to one of a lower grade. Now, however, we shall proceed to reduce the derivation to one of the same grade, but of a lower rank, in order to be able to use the induction hypothesis with respect to  $\rho$ .

The only exception is the first case, 3.121.1, where the mix may at once be altogether eliminated.

In the remaining cases the reduction to derivations of a lower rank is achieved in the following way: The mix is, as it were, moved up one level within the derivation, beyond the inference figure **If**. (Case 3.121.231, for example, illustrates this point particularly well.) To speak more precisely, the left-hand upper sequent of the mix (which from now on will be designated by  $\Pi \rightarrow \Sigma$ ), at

present occurring beside the lower sequent of **If**, is instead written next to the upper sequents of **If**. These now become upper sequents of new mixes. The lower sequents of these mixes are now used as upper sequents of a new inference figure that takes the place of **If**. This new inference figure takes us back either directly, or after having added further inference figures, to the original endsequent. Each new mix obviously has a rank smaller than  $\rho$ , since the left rank remains unchanged and the right rank is diminished by at least 1.

In the strict application of this basic idea special circumstances still arise which make it necessary to distinguish the corresponding cases and to deal with them separately.

3.121.1. Suppose **M** occurs in the antecedent of the left-hand upper sequent of the mix. The end of the derivation runs:

$$\frac{\Pi \rightarrow \Sigma \quad \Delta \rightarrow A}{\Pi, \Delta^* \rightarrow \Sigma^*, A} \text{, thus } \mathbf{M} \text{ occurs in } \Pi.$$

This is transformed into:

$$\frac{\Delta \rightarrow A}{\Pi, \Delta^* \rightarrow \Sigma^*, A} \text{ possibly several thinnings,} \\ \text{contractions and interchanges.}$$

3.121.2. Suppose **M** does not occur in the antecedent of the left-hand upper sequent of the mix. (This assumption is used for the first time in 3.121.222.)

3.121.21. Suppose **If** is a thinning, contraction, or interchange in the antecedent. Then the end of the derivation runs:

$$\frac{\Pi \rightarrow \Sigma \quad \frac{\Psi \rightarrow \theta \quad \text{If}}{\Sigma \rightarrow \theta} \text{ mix.}}{\Pi, \Sigma^* \rightarrow \Sigma^*, \theta}$$

This is transformed into:

$$\frac{\Pi \rightarrow \Sigma \quad \Psi \rightarrow \theta}{\Pi, \Psi^* \rightarrow \Sigma^*, \theta} \text{ mix} \\ \frac{\Psi^*, \Pi \rightarrow \Sigma^*, \theta}{\Sigma^*, \Pi \rightarrow \Sigma^*, \theta} \text{ possibly several interchanges} \\ \frac{\Sigma^*, \Pi \rightarrow \Sigma^*, \theta}{\Pi, \Sigma^* \rightarrow \Sigma^*, \theta} \S \\ \text{possibly several interchanges}$$

The inference figure marked § is of the same kind as **If**, in so far as the  $S$ -formulae designated in the schema of **If** (in 1.21) by **D** and **E**, were not equal to **M**. If **D** or **E** is equal to **M**, we have an identical inference figure ( $\Psi^*$  equals  $\Sigma^*$ ).

The derivation for the lower sequent of the new mix has the same left rank as the old derivation, whereas its right rank is lower by 1. Thus the mix may be completely eliminated by virtue of the induction hypothesis.

3.121.22. Suppose **If** is an inference figure with *one* upper sequent, but not containing a thinning, contraction, or interchange in the antecedent. Then the end of the derivation runs:

$$\frac{\Pi \rightarrow \Sigma \quad \frac{\Psi, \Gamma \rightarrow \Omega_1 \quad \mathbf{If}}{\Xi, \Gamma \rightarrow \Omega_2} \text{mix.}}{\Pi, \Xi^*, \Gamma^* \rightarrow \Sigma^*, \Omega_2}$$

Here we have comprised in  $\Gamma$  the same  $S$ -formulae that are designated by  $\Gamma$  in the schema of the inference figure (1.21, 1.22). Hence  $\Psi$  may be empty or consist of a side formula of the inference figure, and  $\Xi$  may be empty or consist of the principal formula of the inference figure.

First of all, the end of the derivation is transformed into:

$$\frac{\Pi \rightarrow \Sigma \quad \Psi, \Gamma \rightarrow \Omega_1 \text{ mix}}{\Pi, \Psi^*, \Gamma^* \rightarrow \Sigma^*, \Omega_1} \text{ possibly several inter-} \\ \frac{\Psi, \Gamma^*, \Pi \rightarrow \Sigma^*, \Omega_1}{\Xi, \Gamma^*, \Pi \rightarrow \Sigma^*, \Omega_2} \text{ changes and thinnings}$$

The lowest inference is obviously an inference figure of the same kind as **If** (taking  $\Gamma^*, \Pi$  as the  $\Gamma$  of the inference figure and including  $\Sigma^*$  in the  $\Theta$  of the inference figure).

We must only be careful not to violate the restrictions on variables (if **If** is a  $\forall$ -IS or  $\exists$ -IA): Any such violation is precluded by 3.101, which entails that a proper variable that may have occurred in **If** cannot have occurred in  $\Pi$  and  $\Sigma$ .

The mix may be eliminated from the new derivation by virtue of the induction hypothesis.

We therefore obtain a derivation with no mix and which is terminated by the following inference figure:

$$\frac{\Psi, \Gamma^*, \Pi \rightarrow \Sigma^*, \Omega_1}{\Xi, \Gamma^*, \Pi \rightarrow \Sigma^*, \Omega_2}$$

In general, the endsequent is not yet of the form aimed at.

Hence we proceed as follows:

3.121.221. Suppose  $\Xi$  does not contain **M**.

In that case we perform a number of interchanges, if necessary, and obtain the endsequent of the original derivation.

3.121.222. Suppose  $\Xi$  contains **M**. Then  $\Xi$  is the principal formula of **If** and is equal to **M**. We then add:

$$\frac{\Pi \rightarrow \Sigma \quad \mathbf{M}, \Gamma^*, \Pi \rightarrow \Sigma^*, \Omega_2 \text{ mix}}{\Pi, \Gamma^*, \Pi^* \rightarrow \Sigma^*, \Sigma^*, \Omega_2} \text{ possibly several} \\ \frac{\Pi, \Gamma^* \rightarrow \Sigma^*, \Omega_2}{\Pi, \Gamma^* \rightarrow \Sigma^*, \Omega_2} \text{ contractions and} \\ \text{interchanges}$$

Once again, this is the endsequent of the original derivation.

(Above  $\Pi \rightarrow \Sigma$  we once again write the derivation associated with it.)

Thus we have another mix in the derivation. The left rank of our derivation is the same as that of the original derivation. The right rank is now equal to 1. This is so because directly above the right-hand upper sequent occurs the sequent.

$$\Psi, \Gamma^*, \Pi \rightarrow \Sigma^*, \Omega_1$$

**M** no longer occurs in its antecedent, for  $\Gamma^*$  does not contain **M**, nor does  $\Pi$ , because of 3.121.2; and  $\Psi$  contains at most one *side formula* of **If**, which cannot be equal to **M**, since the *principal formula* of **If** is equal to **M**.

Hence this mix, too, may be eliminated by virtue of the induction hypothesis.

3.121.23. Suppose **If** is an inference figure with *two* upper sequents, i.e., an  $\&$ -IS,  $\vee$ -IA, or a  $\supset$ -IA.

(In view of the application to intuitionist logic (3.2) we shall deal with each possibility in greater detail than would be necessary for the classical case.)

3.121.231. Suppose **If** is an  $\&$ -IS.

Then the end of the derivation runs:

$$\frac{\Gamma \rightarrow \Theta, \mathbf{A} \quad \Gamma \rightarrow \Theta, \mathbf{B}}{\Pi \rightarrow \Sigma \quad \Gamma \rightarrow \Theta, \mathbf{A} \& \mathbf{B}} \&\text{-IS} \\ \frac{\Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, \mathbf{A} \& \mathbf{B}}{\Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, \mathbf{A} \& \mathbf{B}} \text{mix}$$

(**M** occurs in  $\Gamma$ .) This is transformed into:

$$\frac{\Pi \rightarrow \Sigma \quad \Gamma \rightarrow \Theta, \mathbf{A} \text{ mix} \quad \Pi \rightarrow \Sigma \quad \Gamma \rightarrow \Theta, \mathbf{B} \text{ mix}}{\Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, \mathbf{A} \text{ mix} \quad \Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, \mathbf{B} \text{ mix}} \&\text{-IS} \\ \frac{\Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, \mathbf{A} \& \mathbf{B}}{\Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, \mathbf{A} \& \mathbf{B}}$$

Both mixes may be eliminated by virtue of the induction hypothesis.

3.121.232. Suppose **If** is a  $\vee$ -IA.

Then the end of the derivation runs:

$$\frac{\mathbf{A}, \Gamma \rightarrow \Theta \quad \mathbf{B}, \Gamma \rightarrow \Theta}{\Pi \rightarrow \Sigma \quad \mathbf{A} \vee \mathbf{B}, \Gamma \rightarrow \Theta} \vee\text{-IA} \\ \frac{\Pi, (\mathbf{A} \vee \mathbf{B})^*, \Gamma^* \rightarrow \Sigma^*, \Theta \text{ mix}}{\Pi, (\mathbf{A} \vee \mathbf{B})^*, \Gamma^* \rightarrow \Sigma^*, \Theta} \text{mix}$$

( $(A \vee B)^*$  stands either for  $A \vee B$  or for nothing according as  $A \vee B$  is unequal or equal to  $M$ .)

$M$  certainly occurs in  $\Gamma$ . (For otherwise  $M$  would be equal to  $A \vee B$ , and the right rank would be equal to 1 contrary to 3.121.)

To begin with, we transform the end of the derivation into:

$$\frac{\frac{\frac{\Pi \rightarrow \Sigma \quad A, \Gamma \rightarrow \Theta}{\Pi, A^*, \Gamma^* \rightarrow \Sigma^*, \Theta} \text{ mix}}{A, \Pi, \Gamma^* \rightarrow \Sigma^*, \Theta} \text{ possibly several interchanges and thinnings}}{\frac{\frac{\Pi \rightarrow \Sigma \quad B, \Gamma \rightarrow \Theta}{\Pi, B^*, \Gamma^* \rightarrow \Sigma^*, \Theta} \text{ mix}}{B, \Pi, \Gamma^* \rightarrow \Sigma^*, \Theta} \text{ possibly several interchanges and thinnings}}{A \vee B, \Pi, \Gamma^* \rightarrow \Sigma^*, \Theta} \vee -IA.$$

Both mixes may be eliminated by virtue of the induction hypothesis.

From here on the procedure is the same as that in 3.121.221 and 3.121.222, i.e., we distinguish two cases according as  $A \vee B$  is unequal or equal to  $M$ . In the first case we may have to add several interchanges to obtain the endsequent of the original derivation; in the second case we add a mix with  $\Pi \rightarrow \Sigma$  for its left-hand upper sequent, and thus once again obtain the endsequent of the original derivation by going on to perform a number of contractions and interchanges, if necessary. The mix concerned may be eliminated, since the associated right rank equals 1. (All this as in 3.121.222.)

3.121.233. Suppose  $If$  is a  $\supset -IA$ .

Then the end of the derivation runs:

$$\frac{\frac{\Gamma \rightarrow \Theta, A \quad B, \Delta \rightarrow A}{A \supset B, \Gamma, \Delta \rightarrow \Theta, A} \supset -IA}{\Pi, (A \supset B)^*, \Gamma^*, \Delta^* \rightarrow \Sigma^*, \Theta, A} \text{ mix}$$

3.121.233.1. Suppose  $M$  occurs in  $\Gamma$  and  $\Delta$ .

In that case we begin by transforming the derivation into:

$$\frac{\frac{\frac{\Pi \rightarrow \Sigma \quad B, \Delta \rightarrow A}{\Pi, B^*, \Delta^* \rightarrow \Sigma^*, A} \text{ mix}}{B, \Pi, \Delta^* \rightarrow \Sigma^*, A} \text{ possibly several interchanges and thinnings}}{\frac{\Pi \rightarrow \Sigma \quad \Gamma \rightarrow \Theta, A}{\Pi, \Gamma^* \rightarrow \Sigma^*, \Theta, A} \text{ mix}}{A \supset B, \Pi, \Gamma^*, \Pi, \Delta^* \rightarrow \Sigma^*, \Theta, \Sigma^*, A} \supset -IA$$

Both mixes may be eliminated by virtue of the induction hypothesis. Then we proceed as in 3.121.221 and 3.121.222. (All that may happen in the first case is that beside interchanges a number of contractions become necessary.)

3.121.233.2. Suppose  $M$  does not occur in both  $\Gamma$  and  $\Delta$  simultaneously.  $M$  must occur in either  $\Gamma$  or  $\Delta$  because of 3.121. Consider the case of  $M$  occurring in  $\Delta$  but not in  $\Gamma$ . The second case is treated analogously.

The end of the derivation is transformed into:

$$\frac{\frac{\frac{\Pi \rightarrow \Sigma \quad B, \Delta \rightarrow A}{\Pi, B^*, \Delta^* \rightarrow \Sigma^*, A} \text{ mix}}{\Gamma \rightarrow \Theta, A \quad B, \Pi, \Delta^* \rightarrow \Sigma^*, A} \text{ possibly several interchanges and thinnings}}{A \supset B, \Gamma, \Pi, \Delta^* \rightarrow \Theta, \Sigma^*, A} \supset -IA$$

The mix may be eliminated by virtue of the induction hypothesis. We then proceed as in 3.121.221 and 3.121.222. (In the second case, where  $A \supset B$  is equal to  $M$ , the right rank belonging to the new mix equals 1 as always, since  $M$  does not occur in  $B, \Pi, \Delta^*$  for the usual reason, nor does it occur in  $\Gamma$  according to the assumption of the case under consideration.)

3.122. Suppose the right rank is equal to 1. In that case the left rank is greater than 1.

This case is, in essence, treated symmetrically to 3.121. Special attention is required only for those inference figures with no symmetric counterpart, viz., the  $\supset -IS$  and the  $\supset -IA$ .

The inference figures  $If$  with one upper sequent were incorporated, in 3.121.22, in the general schema:

$$\frac{\Psi, \Gamma \rightarrow \Omega_1}{\Sigma, \Gamma \rightarrow \Omega_2}$$

The symmetric schema runs:

$$\frac{\Omega_1 \rightarrow \Gamma, \Psi}{\Omega_2 \rightarrow \Gamma, \Sigma}$$

which also includes a  $\supset -IS$  without any further change. ( $\Gamma$  here represents the formulae designated by  $\Theta$  in the schemata 1.21, 1.22.)

3.112.1. On the other hand, the case, where the inference figure  $If$  is a  $\supset -IA$ , must be treated separately. Although this treatment will seem very similar to that in 3.121.233, it is not entirely symmetric.

Thus the end of the derivation runs:

$$\frac{\frac{\Gamma \rightarrow \Theta, A \quad B, \Delta \rightarrow A}{A \supset B, \Gamma, \Delta \rightarrow \Theta, A} \supset -IA}{\frac{A \supset B, \Gamma, \Delta \rightarrow \Theta, A \quad \Sigma \rightarrow \Pi}{A \supset B, \Gamma, \Delta, \Sigma^* \rightarrow \Theta^*, A^*, \Pi} \text{ mix}}$$

3.122.11. Suppose  $M$  occurs both in  $\Theta$  and  $A$ . In that case we transform the end of the derivation into:

$$\begin{array}{c}
 \frac{\Gamma \rightarrow \Theta, A \quad \Sigma \rightarrow \Pi}{\Gamma, \Sigma^* \rightarrow \Theta^*, A^*, \Pi} \text{mix} \\
 \text{possibly several} \\
 \text{interchanges and} \\
 \text{thinnings} \\
 \frac{\Gamma \rightarrow \Theta, A \quad \Sigma \rightarrow \Pi}{\Gamma, \Sigma^* \rightarrow \Theta^*, A^*, \Pi} \text{mix} \\
 \frac{B, \Delta \rightarrow A \quad \Sigma \rightarrow \Pi}{B, \Delta, \Sigma^* \rightarrow A^*, \Pi} \text{mix} \\
 \frac{A \supset B, \Gamma, \Sigma^*, \Delta, \Sigma^* \rightarrow \Theta^*, \Pi, A^*, \Pi}{A \supset B, \Gamma, \Delta, \Sigma^* \rightarrow \Theta^*, A^*, \Pi} \text{possibly several} \\
 \text{contractions and} \\
 \text{interchanges.}
 \end{array}$$

Both mixes may be eliminated by virtue of the induction hypothesis.

3.122.12. Suppose  $M$  does not occur in both  $\Theta$  and  $A$  simultaneously. It must occur in one of them. We consider the case of  $M$  occurring in  $A$  but not in  $\Theta$ ; the alternative case is completely analogous. We transform the end of the derivation into:

$$\frac{\Gamma \rightarrow \Theta, A \quad \frac{B, \Delta \rightarrow A \quad \Sigma \rightarrow \Pi}{B, \Delta, \Sigma^* \rightarrow A^*, \Pi} \text{mix}}{A \supset B, \Gamma, \Delta, \Sigma^* \rightarrow \Theta, A^*, \Pi} \supset -IA$$

The mix may be eliminated by virtue of the induction hypothesis.

### 3.2. Proof of the Hauptsatz for LJ-derivations.

In order to transform an LJ-derivation into an LJ-derivation without cuts, we apply exactly the same procedure as for LK-derivations.

Since an LJ-derivation is a special case of an LK-derivation, it is clear that the transformation can be carried out. We have only to convince ourselves that with every transformation step an LJ-derivation becomes another LJ-derivation, i.e., that the  $D$ -sequents of the transformed derivation do not contain more than one  $S$ -formula in the succedent, given that this was the case before.

We therefore examine each step of the transformation from that point of view.

3.21. Replacement of cuts by mixes. An LJ-cut runs:

$$\frac{\Gamma \rightarrow D \quad D, \Delta \rightarrow A}{\Gamma, \Delta \rightarrow A},$$

where  $A$  contains at most one  $S$ -formula. We transform this cut into:

$$\frac{\Gamma \rightarrow D \quad D, \Delta \rightarrow A}{\Gamma, \Delta^* \rightarrow A} \text{mix} \\
 \frac{\Gamma, \Delta^* \rightarrow A}{\Gamma, \Delta \rightarrow A} \text{possibly several interchanges} \\
 \text{and thinnings in the ante-} \\
 \text{cedent.}$$

This replacement gives us a new LJ-derivation.

3.22. By replacing the free object variable (3.10) we trivially get another LJ-derivation from a previous one.

3.23. The transformation proper (3.11 and 3.12).

We have to show for each of the cases 3.111 to 3.122.12 that the given transformations do not introduce any sequents with more than one  $S$ -formula in the succedent.

3.231. Let us begin with the cases 3.11:

In the cases 3.111, 3.113.1, 3.113.31, 3.113.35 and 3.113.36, only such formulae occur in each succedent of the sequent of a new derivation as already occurred in the succedent of the sequent of the original derivation.

Essentially the same applies in 1.113.33. The only difference is an additional substitution of free object variables, which does not, of course, alter the number of succedent formulae of a sequent.

Cases 3.112, 3.113.32, and 3.113.34 were dealt with symmetrically to cases 3.111, 3.113.1, 3.113.31, and 3.113.33, i.e., in order to get one case from another, we read the schemata from right to left instead of from left to right (as well as changing logical symbols, a process which is here of no consequence). Hence in the antecedent of one case we get precisely the same as in the succedent of another. For the antecedents of cases 3.111, 3.113.1, 3.113.31 and 3.113.33, the same applies as for the succedents, viz., in every antecedent of a sequent of the new derivation only such formulae occur as already occurred in an antecedent of a sequent of the original derivation.

This disposes of all symmetric cases: 3.112, 3.113.2, 3.113.32 and 3.113.34.

3.23. Now let us look at the cases, 3.12:

3.232.1. For the cases 3.121 it holds generally that  $\Sigma^*$  is empty, since in  $\Pi \rightarrow \Sigma$ ,  $\Sigma$  must contain only one formula, and that formula must be equal to  $M$ .

It is now obvious that in every succedent of a sequent only such formulae occur as already occurred in the succedent of a sequent of the original derivation.

3.232.2. In the cases 3.122 it is somewhat more difficult to see that from an LJ-derivation we always get another LJ-derivation. We must direct our attention, as we have already done in considering symmetric cases, to the antecedents in the schemata 3.121.

At this point we distinguish two further sub-cases:

3.232.21. The case which is symmetric to 3.121.1 is trivial, since in every antecedent of a sequent of a new derivation (in case 3.121.1) only such formulae occur as already occurred in an antecedent of a sequent of the original derivation.

3.232.22. In the cases that are symmetric to 3.121.2, the mix in the end of the derivation runs:

$$\frac{\Omega \rightarrow \mathbf{M} \quad \Sigma \rightarrow \Pi}{\Omega, \Sigma^* \rightarrow \Pi}$$

where  $\Pi$  contains at most *one*  $S$ -formula, and where  $\Omega \rightarrow \mathbf{M}$  is the lower sequent of an  $LJ$ -inference figure in which at least one upper sequent contains  $\mathbf{M}$  as a succedent formula.

If we now look at the inference figure schemata 1.21, 1.22, it becomes easily apparent that such an inference figure can only be a thinning, contraction, or interchange in the antecedent, or a  $\vee$ - $IA$ , a  $\&$ - $IA$ , a  $\exists$ - $IA$ , a  $\forall$ - $IA$ , and a  $\supset$ - $IA$ . Let us disregard for the moment the  $\vee$ - $IA$  and the  $\supset$ - $IA$ . Then all the possibilities enumerated above fall within the case symmetric to 3.121.22, where

both  $\Psi$  and  $\Sigma$  always remain empty. ( $\Gamma$  corresponds to the  $\Theta$  of the inference figure.) Thus we have the case which is symmetric to 3.121.221. Furthermore,  $\Gamma$  is equal to  $\mathbf{M}$ , i.e.,  $\Gamma^*$  is empty, and  $\Pi$  contains at most one formula. Hence in the new derivation there never in fact occurs more than one formula in the succedent of a sequent.

The case of a  $\vee$ - $IA$  is symmetric to 3.121.231. Again,  $\Gamma$  is equal to  $\mathbf{M}$ ,  $\Gamma^*$  is empty, and  $\Pi$  contains at most one formula; all is thus in order.

There now remains the case of a  $\supset$ - $IA$ , i.e., 3.122.1. In an  $LJ$ - $\supset$ - $IA$ , the  $\Theta$  of the schema (1.22) is empty. Thus we have the case set out under 3.122.12.  $\Lambda^*$  is also empty, and  $\Pi$  contains at most one formula, which means that here, too, we again obtain an  $LJ$ -derivation from an  $LJ$ -derivation.

## GLOSSARY

*All-Zeichen*—universal quantifier  
*Annahmeformel*—assumption formula  
*Antezedenz*—antecedent  
*Äußerstes Zeichen*—terminal symbol  
*Eigenvariable*—proper variable  
*Es-gibt-Zeichen*—existential quantifier  
*Faden*—branch  
*Folgt-Zeichen*—implication symbol  
*Grad*—grade  
*Grundformel*—basic formula  
*Hauptformel*—principal formula  
*Hauptsatz*—*Hauptsatz*  
*Herleitung*—derivation  
*Hilfssatz*—lemma  
*Inhaltlicher Sinn*—intuitive sense  
*Logische-Zeichen-Schlußfigur*—operational inference figure  
*Mischformel*—mix formula  
*Mischung*—mix  
*Mittelungszeichen*—syntactic variable  
*Nebenformel*—aide formula  
*Nicht-Zeichen*—negation symbol

*Oberformel*—upper formula  
*Obersequenz*—upper sequent  
*Oder-Zeichen*—disjunction symbol  
*Rang*—rank  
*Satz*—theorem  
*Schließen*—deduction  
*Schluß*—inference  
*Schnitt*—cut  
*Sequenz*—sequent  
*Spiegelbildlich*—dual  
*Stammbaumform*—tree form  
*Struktur-Schlußfigur*—structural inference figure  
*Sukzedenz*—succedent  
*Teilformel*—subformula  
*Und-Zeichen*—conjunction symbol  
*Untersequenz*—lower sequent  
*Verdünnung*—thinning  
*Vertauschung*—interchange  
*Zeichen für Bestimmtes*—constant symbol  
*Zusammenziehung*—contraction

### III. UNDERSTANDING A PRIMITIVE SOCIETY

PETER WINCH

THIS essay will pursue further some questions raised in my book, *The Idea of a Social Science*.<sup>1</sup> That book was a general discussion of what is involved in the understanding of human social life. I shall here be concerned more specifically with certain issues connected with social anthropology. In the first part I raise certain difficulties about Professor E. E. Evans-Pritchard's approach in his classic, *Witchcraft, Oracles and Magic among the Azande*.<sup>2</sup> In the second part, I attempt to refute some criticisms recently made by Mr. Alasdair MacIntyre of Evans-Pritchard and myself, to criticize in their turn MacIntyre's positive remarks, and to offer some further reflections of my own on the concept of learning from the study of a primitive society.

#### I. THE REALITY OF MAGIC

Like many other primitive people, the African Azande hold beliefs that we cannot possibly share and engage in practices which it is peculiarly difficult for us to comprehend. They believe that certain of their members are witches, exercising a malignant occult influence on the lives of their fellows. They engage in rites to counteract witchcraft; they consult oracles and use magic medicines to protect themselves from harm.

An anthropologist studying such a people wishes to make those beliefs and practices intelligible to himself and his readers. This means presenting an account of them that will somehow satisfy the criteria of rationality demanded by the culture to which he and his readers belong: a culture whose conception of rationality is deeply affected by the achievements and methods of the sciences, and one which treats such things as a belief in magic or the practice of consulting oracles as almost a paradigm of the irrational. The strains inherent in this situation are very likely to lead the anthropologist to

adopt the following posture: *We* know that Zande beliefs in the influence of witchcraft, the efficacy of magic medicines, the role of oracles in revealing what is going on and what is going to happen, are mistaken, illusory. Scientific methods of investigation have shown conclusively that there are no relations of cause and effect such as are implied by these beliefs and practices. All we can do then is to show how such a system of mistaken beliefs and inefficacious practices can maintain itself in the face of objections that seem to us so obvious.<sup>3</sup>

Now although Evans-Pritchard goes a very great deal further than most of his predecessors in trying to present the sense of the institutions he is discussing as it presents itself to the Azande themselves, still, the last paragraph does, I believe, pretty fairly describe the attitude he himself took at the time of writing this book. There is more than one remark to the effect that "obviously there are no witches"; and he writes of the difficulty he found, during his field work with the Azande, in shaking off the "unreason" on which Zande life is based and returning to a clear view of how things really are. This attitude is not an unsophisticated one but is based on a philosophical position ably developed in a series of papers published in the 1930's in the unhappily rather inaccessible *Bulletin of the Faculty of Arts* of the University of Egypt. Arguing against Lévy-Bruhl, Evans-Pritchard here rejects the idea that the scientific understanding of causes and effects which leads us to reject magical ideas is evidence of any superior intelligence on our part. Our scientific approach, he points out, is as much a function of our culture as is the magical approach of the "savage" a function of his.

The fact that we attribute rain to meteorological causes alone while savages believe that Gods or ghosts or magic can influence the rainfall is no evidence that our brains function differently from their brains. It does not show that we "think more logically" than

<sup>1</sup> London and New York (Routledge & Kegan Paul; Humanities Press), 1958.

<sup>2</sup> Oxford (Oxford University Press), 1937.

<sup>3</sup> At this point the anthropologist is very likely to start speaking of the "social function" of the institution under examination. There are many important questions that should be raised about functional explanations and their relations to the issues discussed in this essay; but these questions cannot be pursued further here.

savages, at least not if this expression suggests some kind of hereditary psychic superiority. It is no sign of superior intelligence on my part that I attribute rain to physical causes. I did not come to this conclusion myself by observation and inference and have, in fact, little knowledge of the meteorological processes that lead to rain. I merely accept what everybody else in my society accepts, namely that rain is due to natural causes. This particular idea formed part of my culture long before I was born into it and little more was required of me than sufficient linguistic ability to learn it. Likewise a savage who believes that under suitable natural and ritual conditions the rainfall can be influenced by use of appropriate magic is not on account of this belief to be considered of inferior intelligence. He did not build up this belief from his own observations and inferences but adopted it in the same way as he adopted the rest of his cultural heritage, namely, by being born into it. He and I are both thinking in patterns of thought provided for us by the societies in which we live.

It would be absurd to say that the savage is thinking mystically and that we are thinking scientifically about rainfall. In either case like mental processes are involved and, moreover, the content of thought is similarly derived. But we can say that the social content of our thought about rainfall is scientific, is in accord with objective facts, whereas the social content of savage thought about rainfall is unscientific since it is not in accord with reality and may also be mystical where it assumes the existence of supra-sensible forces.<sup>4</sup>

In a subsequent article on Pareto, Evans-Pritchard distinguishes between "logical" and "scientific."

Scientific notions are those which accord with objective reality both with regard to the validity of their premisses and to the inferences drawn from their propositions. . . . Logical notions are those in which according to the rules of thought inferences would be true were the premisses true, the truth of the premisses being irrelevant. . . .

A pot has broken during firing. This is probably due to grit. Let us examine the pot and see if this is the cause. That is logical and scientific thought. Sickness is due to witchcraft. A man is sick. Let us consult the oracles to discover who is the witch responsible. That is logical and unscientific thought.<sup>5</sup>

I think that Evans-Pritchard is right in a great deal of what he says here, but wrong, and crucially wrong, in his attempt to characterize the scientific in terms of that which is "in accord with objective

reality." Despite differences of emphasis and phraseology, Evans-Pritchard is in fact hereby put into the same metaphysical camp as Pareto: for both of them the conception of "reality" must be regarded as intelligible and applicable *outside* the context of scientific reasoning itself, since it is that to which scientific notions do, and unscientific notions do not, have a relation. Evans-Pritchard, although he emphasizes that a member of scientific culture has a different conception of reality from that of a Zande believer in magic, wants to go beyond merely registering this fact and making the differences explicit, and to say, finally, that the scientific conception agrees with what reality actually is like, whereas the magical conception does not.

It would be easy, at this point, to say simply that the difficulty arises from the use of the unwieldy and misleadingly comprehensive expression "agreement with reality"; and in a sense this is true. But we should not lose sight of the fact that the idea that men's ideas and beliefs must be checkable by reference to something independent—some reality—is an important one. To abandon it is to plunge straight into an extreme Protagorean relativism, with all the paradoxes that involves. On the other hand great care is certainly necessary in fixing the precise role that this conception of the independently real does play in men's thought. There are two related points that I should like to make about it at this stage.

In the first place we should notice that the check of the independently real is not peculiar to science. The trouble is that the fascination science has for us makes it easy for us to adopt its scientific form as a paradigm against which to measure the intellectual respectability of other modes of discourse. Consider what God says to Job out of the whirlwind: "Who is this that darkeneth counsel by words without knowledge? . . . Where wast thou when I laid the foundations of the earth? declare, if thou hast understanding. Who hath laid the measures thereof, if thou knowest? or who hath stretched the line upon it. . . . Shall he that contendeth with the Almighty instruct him? he that reproveth God, let him answer it." Job is taken to task for having gone astray by having lost sight of the reality of God; this does not, of course, mean that Job has made any sort of theoretical mistake, which could be put right, perhaps, by means of an

<sup>4</sup> E. E. Evans-Pritchard, "Lévy-Bruhl's Theory of Primitive Mentality," *Bulletin of the Faculty of Arts, University of Egypt*, 1934.

<sup>5</sup> "Science and Sentiment," *Bulletin of the Faculty of Arts, ibid.*, 1935.

experiment.<sup>6</sup> God's reality is certainly independent of what any man may care to think, but what that reality amounts to can only be seen from the religious tradition in which the concept of God is used, and this use is very unlike the use of scientific concepts, say of theoretical entities. The point is that it is *within* the religious use of language that the conception of God's reality has its place, though, I repeat, this does not mean that it is at the mercy of what anyone cares to say; if this were so, God would have no reality.

My second point follows from the first. Reality is not what gives language sense. What is real and what is unreal shows itself *in* the sense that language has. Further, both the distinction between the real and the unreal and the concept of agreement with reality themselves belong to our language. I will not say that they are concepts of the language like any other, since it is clear that they occupy a commanding, and in a sense a limiting, position there. We can imagine a language with no concept of, say, wetness, but hardly one in which there is no way of distinguishing the real from the unreal. Nevertheless we could not in fact distinguish the real from the unreal without understanding the way this distinction operates in the language. If then we wish to understand the significance of these concepts, we must examine the use they actually do have—in the language.

Evans-Pritchard, on the contrary, is trying to work with a conception of reality which is *not* determined by its actual use in language. He wants something against which that use can itself be appraised. But this is not possible; and no more possible in the case of scientific discourse than it is in any other. We may ask whether a particular scientific hypothesis agrees with reality and test this by observation and experiment. Given the experimental methods, and the established use of the theoretical terms entering into the hypothesis, then the question whether it holds or not is settled by reference to something independent of what I, or anybody else, care to think. But the general nature of the data revealed by the experiment can only be specified in terms of criteria built into the methods of experiment employed and these, in turn, make sense only to someone who is conversant with the kind of scientific activity within which they are employed. A scientific illiterate, asked to describe the results of an experiment which he "observes" in an advanced physics laboratory, could not do so

in terms relevant to the hypothesis being tested; and it is really only in such terms that we can sensibly speak of the "results of the experiment" at all. What Evans-Pritchard wants to be able to say is that the criteria applied in scientific experimentation constitute a true link between our ideas and an independent reality, whereas those characteristic of other systems of thought—in particular, magical methods of thought—do not. It is evident that the expressions "true link" and "independent reality" in the previous sentence cannot themselves be explained by reference to the scientific universe of discourse, as this would beg the question. We have then to ask how, by reference to what established universe of discourse, the use of those expressions is to be explained; and it is clear that Evans-Pritchard has not answered this question.

Two questions arise out of what I have been saying. First, is it in fact the case that a primitive system of magic, like that of the Azande, constitutes a coherent universe of discourse like science, in terms of which an intelligible conception of reality and clear ways of deciding what beliefs are and are not in agreement with this reality can be discerned? Second, what are we to make of the possibility of understanding primitive social institutions, like Zande magic, if the situation is as I have outlined? I do not claim to be able to give a satisfactory answer to the second question. It raises some very important and fundamental issues about the nature of human social life, which require conceptions different from, and harder to elucidate, than those I have hitherto introduced. I shall offer some tentative remarks about these issues in the second part of this essay. At present I shall address myself to the first question.

It ought to be remarked here that an affirmative answer to my first question would not commit me to accepting as rational all beliefs couched in magical concepts or all procedures practiced in the name of such beliefs. This is no more necessary than is the corresponding proposition that all procedures "justified" in the name of science are immune from rational criticism. A remark of Collingwood's is apposite here:

Savages are no more exempt from human folly than civilized men, and are no doubt equally liable to the error of thinking that they, or the persons they regard as their superiors, can do what in fact cannot be done. But this error is not the essence of magic; it is a perversion of magic. And we should be careful how we

<sup>6</sup> Indeed, one way of expressing the point of the story of Job is to say that in it Job is shown as going astray by being induced to make the reality and goodness of God contingent on what happens.

attribute it to the people we call savages, who will one day rise up and testify against us.<sup>7</sup>

It is important to distinguish a system of magical beliefs and practices like that of the Azande, which is one of the principal foundations of their whole social life and, on the other hand, magical beliefs that might be held, and magical rites that might be practiced, by persons belonging to our own culture. These have to be understood rather differently. Evans-Pritchard is himself alluding to the difference in the following passage: "When a Zande speaks of witchcraft he does not speak of it as we speak of the weird witchcraft of our own history. Witchcraft is to him a commonplace happening and he seldom passes a day without mentioning it. . . . To us witchcraft is something which haunted and disgusted our credulous forefathers. But the Zande expects to come across witchcraft at any time of the day or night. He would be just as surprised if he were not brought into daily contact with it as we would be if confronted by its appearance. To him there is nothing miraculous about it."<sup>8</sup>

The difference is not merely one of degree of familiarity, however, although, perhaps, even this has more importance than might at first appear. Concepts of witchcraft and magic in our culture, at least since the advent of Christianity, have been parasitic on, and a perversion of other orthodox concepts, both religious and, increasingly, scientific. To take an obvious example, you could not understand what was involved in conducting a Black Mass, unless you were familiar with the conduct of a proper Mass and, therefore, with the whole complex of religious ideas from which the Mass draws its sense. Neither would you understand the relation between these without taking account of the fact that the Black practices are rejected as *irrational* (in the sense proper to religion) in the system of beliefs on which these practices are thus parasitic. Perhaps a similar relation holds between the contemporary practice of astrology and astronomy and technology. It is impossible to keep a discussion of the rationality of Black Magic or of astrology within the bounds of concepts peculiar to them; they have an essential reference to something outside themselves. The position is like that which Socrates, in Plato's *Gorgias*, showed to be true of the Sophists' conception of rhetoric: namely, that it is parasitic on rational discourse in

such a way that its irrational character can be shown in terms of this dependence. Hence, when we speak of such practices as "superstitious," "illusory," "irrational," we have the weight of our culture behind us; and this is not just a matter of being on the side of the big battalions, because those beliefs and practices belong to, and derive such sense as they seem to have, from that same culture. This enables us to show that the sense is only apparent, in terms which are culturally relevant.

It is evident that our relation to Zande magic is quite different. If we wish to understand it, we must seek a foothold elsewhere. And while there may well be room for the use of such critical expressions as "superstition" and "irrationality," the kind of rationality with which such terms might be used to point a contrast remains to be elucidated. The remarks I shall make in Part II will have a more positive bearing on this issue. In the rest of this Part, I shall develop in more detail my criticisms of Evans-Pritchard's approach to the Azande.

Early in this book he defines certain categories in terms of which his descriptions of Zande customs are couched.

MYSTICAL NOTIONS . . . are patterns of thought that attribute to phenomena supra-sensible qualities which, or part of which, are not derived from observation or cannot be logically inferred from it, and which they do not possess.<sup>9</sup> COMMON-SENSE NOTIONS . . . attribute to phenomena only what men observe in them or what can logically be inferred from observation. So long as a notion does not assert something which has not been observed, it is not classed as mystical even though it is mistaken on account of incomplete observation. . . . SCIENTIFIC NOTIONS. Science has developed out of common-sense but is far more methodical and has better techniques of observation and reasoning. Common sense uses experience and rules of thumb. Science uses experiment and rules of Logic. . . . *Our body of scientific knowledge and Logic are the sole arbiters of what are mystical, common sense, and scientific notions.* Their judgments are never absolute. RITUAL BEHAVIOUR. Any behaviour that is accounted for by mystical notions. *There is no objective nexus between the behaviour and the event it is intended to cause. Such behaviour is usually intelligible to us only when we know the mystical notions associated with it.* EMPIRICAL BEHAVIOUR. Any behaviour that is accounted for by common-sense notions.<sup>10</sup>

<sup>7</sup> R. G. Collingwood, *Principles of Art* (Oxford, Oxford University Press, Galaxy Books, 1958), p. 67.

<sup>8</sup> *Witchcraft, Oracles and Magic among the Azande*, p. 64.

<sup>9</sup> The italics are mine throughout this quotation.

<sup>10</sup> *Op. cit.*, p. 12.

It will be seen from the phrases which I have italicized that Evans-Pritchard is doing more here than just defining certain terms for his own use. Certain metaphysical claims are embodied in the definitions: identical in substance with the claims embodied in Pareto's way of distinguishing between "logical" and "non-logical" conduct.<sup>11</sup> There is a very clear implication that those who use mystical notions and perform ritual behavior are making some sort of mistake, detectable with the aid of science and logic. I shall now examine more closely some of the institutions described by Evans-Pritchard to determine how far his claims are justified.

*Witchcraft* is a power possessed by certain individuals to harm other individuals by "mystical" means. Its basis is an inherited organic condition, "witchcraft-substance" and it does not involve any special magical ritual or medicine. It is constantly appealed to by Azande when they are afflicted by misfortune, not so as to exclude explanation in terms of natural causes, which Azande are perfectly able to offer themselves within the limits of their not inconsiderable natural knowledge, but so as to supplement such explanations. "Witchcraft explains *why*<sup>12</sup> events are harmful to man and not *how*<sup>13</sup> they happen. A Zande perceives how they happen just as we do. He does not see a witch charge a man, but an elephant. He does not see a witch push over the granary, but termites gnawing away its supports. He does not see a psychical flame igniting thatch, but an ordinary lighted bundle of straw. His perception of how events occur is as clear as our own."<sup>13</sup>

The most important way of detecting the influence of witchcraft and of identifying witches is by the revelations of oracles, of which in turn the most important is the "poison oracle." This name, though convenient, is significantly misleading insofar as, according to Evans-Pritchard, Azande do not have our concept of a poison and do not think of, or behave toward, *benge*—the substance administered in the consultation of the oracle—as we do of and toward poisons. The gathering, preparation, and administering of *benge* is hedged with ritual and strict taboos. At an oracular consultation *benge* is administered to a fowl, while a question is asked in a form permitting a yes or no answer. The fowl's death or survival is specified beforehand as giving the answer "yes" or

"no." The answer is then checked by administering *benge* to another fowl and asking the question the other way round. "Is Prince Ndoruma responsible for placing bad medicines in the roof of my hut? The fowl DIES giving the answer 'Yes.' . . . Did the oracle speak truly when it said that Ndoruma was responsible? The fowl SURVIVES giving the answer 'Yes'." The poison oracle is all-pervasive in Zande life and all steps of any importance in a person's life are settled by reference to it.

A Zande would be utterly lost and bewildered without his oracle. The mainstay of his life would be lacking. It is rather as if an engineer, in our society, were to be asked to build a bridge without mathematical calculation, or a military commander to mount an extensive co-ordinated attack without the use of clocks. These analogies are mine, but a reader may well think that they beg the question at issue. For, he may argue, the Zande practice of consulting the oracle, unlike my technological and military examples, is completely unintelligible and rests on an obvious illusion. I shall now consider this objection.

First I must emphasize that I have so far done little more than note the *fact*, conclusively established by Evans-Pritchard, that the Azande *do* in fact conduct their affairs to their own satisfaction in this way and are at a loss when forced to abandon the practice—when, for instance, they fall into the hands of European courts. It is worth remarking too that Evans-Pritchard himself ran his household in the same way during his field researches and says: "I found this as satisfactory a way of running my home and affairs as any other I know of."

Further, I would ask in my turn: *to whom* is the practice alleged to be unintelligible? Certainly it is difficult for us to understand what the Azande are about when they consult their oracles; but it might seem just as incredible to them that the engineer's motions with his slide rule could have any connection with the stability of his bridge. But this riposte of course misses the intention behind the objection, which was not directed to the question whether anyone in fact understands, or claims to understand, what is going on, but rather whether what is going on actually does make sense: i.e., in itself. And it may seem obvious that Zande beliefs in witchcraft and oracles cannot make any

<sup>11</sup> For further criticism of Pareto see Peter Winch, *The Idea of a Social Science*, pp. 95-111.

<sup>12</sup> Evans-Pritchard's italics.

<sup>13</sup> *Op. cit.*, p. 72.

sense, however satisfied the Azande may be with them.

What criteria have we for saying that something does, or does not, make sense? A partial answer is that a set of beliefs and practices cannot make sense insofar as they involve contradictions. Now it appears that contradictions are bound to arise in at least two ways in the consultation of the oracle. On the one hand two oracular pronouncements may contradict each other; and on the other hand a self-consistent oracular pronouncement may be contradicted by future experience. I shall examine each of these apparent possibilities in turn.

Of course, it does happen often that the oracle first says "yes" and then "no" to the same question. This does not convince a Zande of the futility of the whole operation of consulting oracles: obviously, it cannot, since otherwise the practice could hardly have developed and maintained itself at all. Various explanations may be offered, whose possibility, it is important to notice, is built into the whole network of Zande beliefs and may, therefore, be regarded as belonging to the concept of an oracle. It may be said, for instance, that bad *benge* is being used; that the operator of the oracle is ritually unclean; that the oracle is being itself influenced by witchcraft or sorcery; or it may be that the oracle is showing that the question cannot be answered straightforwardly in its present form, as with "Have you stopped beating your wife yet?" There are various ways in which the behavior of the fowl under the influence of *benge* may be ingeniously interpreted by those wise in the ways of the poison oracle. We might compare this situation perhaps with the interpretation of dreams.

In the other type of case: where an internally consistent oracular revelation is apparently contradicted by subsequent experience, the situation may be dealt with in a similar way, by references to the influence of witchcraft, ritual uncleanness, and so on. But there is another important consideration we must take into account here too. The chief function of oracles is to reveal the presence of "mystical" forces—I use Evans-Pritchard's term without committing myself to his denial that such forces really exist. Now though there are indeed ways of determining whether or not mystical forces are operating, these ways do not correspond to what we understand by "empirical" confirmation or refutation. This indeed is a tautology, since such differences in "confirmatory" procedures

are the main criteria for classifying something as a mystical force in the first place. Here we have one reason why the possibilities of "refutation by experience" are very much fewer than might at first sight be supposed.

There is also another closely connected reason. The spirit in which oracles are consulted is very unlike that in which a scientist makes experiments. Oracular revelations are not treated as hypotheses and, since their sense derives from the way they are treated in their context, they therefore *are not* hypotheses. They are not a matter of intellectual interest but the main way in which Azande decide how they should act. If the oracle reveals that a proposed course of action is fraught with mystical dangers from witchcraft or sorcery, that course of action will not be carried out; and then the question of refutation or confirmation just does not arise. We might say that the revelation has the logical status of an unfulfilled hypothetical, were it not that the context in which this logical term is generally used may again suggest a misleadingly close analogy with scientific hypotheses.

I do not think that Evans-Pritchard would have disagreed with what I have said so far. Indeed, the following comment is on very similar lines:

Azande observe the action of the poison oracle as we observe it, but their observations are always subordinated to their beliefs and are incorporated into their beliefs and made to explain them and justify them. Let the reader consider any argument that would utterly demolish all Zande claims for the power of the oracle. If it were translated into Zande modes of thought it would serve to support their entire structure of belief. For their mystical notions are eminently coherent, being interrelated by a network of logical ties, and are so ordered that they never too crudely contradict sensory experience but, instead, experience seems to justify them. The Zande is immersed in a sea of mystical notions, and if he speaks about his poison oracle he must speak in a mystical idiom.<sup>14</sup>

To locate the point at which the important philosophical issue does arise, I shall offer a parody, composed by changing round one or two expressions in the foregoing quotation.

Europeans observe the action of the poison oracle just as Azande observe it, but their observations are always subordinated to their beliefs and are incorporated into their beliefs and made to explain them and justify them. Let a Zande consider any argument that would utterly refute all European scepticism about the power of the oracle. If it were translated into

<sup>14</sup> *Ibid.*, p. 319.

European modes of thought it would serve to support their entire structure of belief. For their scientific notions are eminently coherent, being interrelated by a network of logical ties, and are so ordered that they never too crudely contradict mystical experience but, instead, experience seems to justify them. The European is immersed in a sea of scientific notions, and if he speaks about the Zande poison oracle he must speak in a scientific idiom.

Perhaps this too would be acceptable to Evans-Pritchard. But it is clear from other remarks in the book to which I have alluded, that at the time of writing it he would have wished to add: and the European is right and the Zande wrong. This addition I regard as illegitimate and my reasons for so thinking take us to the heart of the matter.

It may be illuminating at this point to compare the disagreement between Evans-Pritchard and me to that between the Wittgenstein of the *Philosophical Investigations* and his earlier *alter ego* of the *Tractatus Logico-Philosophicus*. In the *Tractatus* Wittgenstein sought "the general form of propositions": what made propositions possible. He said that this general form is: "This is how things are"; the proposition was an articulated model, consisting of elements standing in a definite relation to each other. The proposition was true when there existed a corresponding arrangement of elements in reality. The proposition was capable of saying something because of the identity of structure, of logical form, in the proposition and in reality.

By the time Wittgenstein composed the *Investigations* he had come to reject the whole idea that there must be a general form of propositions. He emphasized the indefinite number of different uses that language may have and tried to show that these different uses neither need, nor in fact do, all have something in common, in the sense intended in the *Tractatus*. He also tried to show that what counts as "agreement or disagreement with reality" takes on as many different forms as there are different uses of language and cannot, therefore, be taken as given *prior* to the detailed investigation of the use that is in question.

The *Tractatus* contains a remark strikingly like something that Evans-Pritchard says.

*The limits of my language mean the limits of my world. Logic fills the world: the limits of the world are also its limits. We cannot therefore say in logic: This and this there is in the world, and that there is not.*

For that would apparently presuppose that we

exclude certain possibilities, and this cannot be the case since otherwise logic must get outside the limits of the world: that is, if it could consider these limits from the other side also.<sup>15</sup>

Evans-Pritchard discusses the phenomena of belief and scepticism, as they appear in Zande life. There is certainly widespread scepticism about certain things, for instance, about some of the powers claimed by witchdoctors or about the efficacy of certain magic medicines. But, he points out, such scepticism does not begin to overturn the mystical way of thinking, since it is necessarily expressed in terms belonging to that way of thinking.

In this web of belief every strand depends on every other strand, and a Zande cannot get outside its meshes because this is the only world he knows. The web is not an external structure in which he is enclosed. It is the texture of his thought and he cannot think that his thought is wrong.<sup>16</sup>

Wittgenstein and Evans-Pritchard are concerned here with much the same problem, though the difference in the directions from which they approach it is important too. Wittgenstein, at the time of the *Tractatus*, spoke of "language," as if all language is fundamentally of the same kind and must have the same kind of "relation to reality"; but Evans-Pritchard is confronted by two languages which he recognizes as fundamentally different in kind, such that much of what may be expressed in the one has no possible counterpart in the other. One might, therefore, have expected this to lead to a position closer to that of the *Philosophical Investigations* than to that of the *Tractatus*. Evans-Pritchard is not content with elucidating the differences in the two concepts of reality involved; he wants to go further and say: our concept of reality is the correct one, the Azande are mistaken. But the difficulty is to see what "correct" and "mistaken" can mean in this context.

Let me return to the subject of contradictions. I have already noted that many contradictions we might expect to appear in fact do not in the context of Zande thought, where provision is made for avoiding them. But there are some situations of which this does not seem to be true, where what appear to us as obvious contradictions are left where they are, apparently unresolved. Perhaps this may be the foothold we are looking for, from

<sup>15</sup> Wittgenstein, *Tractatus Logico-Philosophicus*, paras. 5.6-5.61.

<sup>16</sup> Evans-Pritchard, *op. cit.*, p. 194.

which we can appraise the "correctness" of the Zande system.<sup>17</sup>

Consider Zande notions about the inheritance of witchcraft. I have spoken so far only of the role of oracles in establishing whether or not someone is a witch. But there is a further and, as we might think, more "direct" method of doing this, namely by post-mortem examination of a suspect's intestines for "witchcraft-substance." This may be arranged by his family after his death in an attempt to clear the family name of the imputation of witchcraft. Evans-Pritchard remarks: "to our minds it appears evident that if a man is proven a witch the whole of his clan are *ipso facto* witches, since the Zande clan is a group of persons related biologically to one another through the male line. Azande see the sense of this argument but they do not accept its conclusions, and it would involve the whole notion of witchcraft in contradiction were they to do so."<sup>18</sup> Contradiction would presumably arise because a few positive results of post-mortem examinations, scattered among all the clans, would very soon prove that everybody was a witch, and a few negative results, scattered among the same clans, would prove that nobody was a witch. Though, in particular situations, individual Azande may avoid personal implications arising out of the presence of witchcraft-substance in deceased relatives, by imputations of bastardy and similar devices, this would not be enough to save the generally contradictory situation I have sketched. Evans-Pritchard comments: "Azande do not perceive the contradiction as we perceive it because they have no theoretical interest in the subject, and those situations in which they express their belief in witchcraft do not force the problem upon them."<sup>19</sup>

It might now appear as though we had clear grounds for speaking of the superior rationality of European over Zande thought, insofar as the latter involves a contradiction which it makes no attempt to remove and does not even recognize: one, however, which is recognizable as such in the context of European ways of thinking. But does Zande thought on this matter really involve a contradiction? It appears from Evans-Pritchard's account that Azande do not press their ways of thinking about witches to a point at which they would be involved in contradictions.

Someone may now want to say that the irrationality of the Azande in relation to witchcraft shows itself in the fact that they do not press their thought about it "to its logical conclusion." To appraise this point we must consider whether the conclusion we are trying to force on them is indeed a logical one; or perhaps better, whether someone who does press this conclusion is being more rational than the Azande, who do not. Some light is thrown on this question by Wittgenstein's discussion of a game,

such that whoever begins can always win by a particular simple trick. But this has not been realized—so it is a game. Now someone draws our attention to it—and it stops being a game.

What turn can I give this, to make it clear to myself?—For I want to say: "and it stops being a game"—not: "and now we see that it wasn't a game."

That means, I want to say, it can also be taken like this: the other man did not *draw our attention* to anything; he taught us a different game in place of our own. But how can the new game have made the old one obsolete? We now see something different, and can no longer naively go on playing.

On the one hand the game consisted in our actions (our play) on the board; and these actions I could perform as well now as before. But on the other hand it was essential to the game that I blindly tried to win; and now I can no longer do that.<sup>20</sup>

There are obviously considerable analogies between Wittgenstein's example and the situation we are considering. But there is an equally important difference. Both Wittgenstein's games: the old one without the trick that enables the starter to win and the new one with the trick, are in an important sense on the same level. They are both *games*, in the form of a contest where the aim of a player is to beat his opponent by the exercise of skill. The new trick makes this situation impossible and this is why it makes the old game obsolete. To be sure, the situation could be saved in a way by introducing a new rule, forbidding the use by the starter of the trick which would ensure his victory. But our intellectual habits are such as to make us unhappy about the artificiality of such a device, rather as logicians have been unhappy about the introduction of a Theory of Types as a device for avoiding Russell's paradoxes. It is noteworthy in my last quotation from Evans-Pritchard, however, that the Azande, when the possibility of

<sup>17</sup> I shall discuss this point in a more general way in Part II.

<sup>18</sup> *Ibid.*, p. 24.

<sup>19</sup> *Ibid.*, p. 25.

<sup>20</sup> L. Wittgenstein, *Remarks on the Foundations of Mathematics*, Pt. II, § 77. Wittgenstein's whole discussion of "contradiction" in mathematics is directly relevant to the point I am discussing.

this contradiction about the inheritance of witchcraft is pointed out to them, do *not* then come to regard their old beliefs about witchcraft as obsolete. "They have no theoretical interest in the subject." This suggests strongly that the context from which the suggestion about the contradiction is made, the context of our scientific culture, is not on the same level as the context in which the beliefs about witchcraft operate. Zande notions of witchcraft do not constitute a theoretical system in terms of which Azande try to gain a quasi-scientific understanding of the world.<sup>21</sup> This in its turn suggests that it is the European, obsessed with pressing Zande thought where it would not naturally go—to a contradiction—who is guilty of misunderstanding, not the Zande. The European is in fact committing a category-mistake.

Something else is also suggested by this discussion: the forms in which rationality expresses itself in the culture of a human society cannot be elucidated *simply* in terms of the logical coherence of the rules according to which activities are carried out in that society. For, as we have seen, there comes a point where we are not even in a position to determine what is and what is not coherent in such a context of rules, without raising questions about the point which following those rules has in the society. No doubt it was a realization of this fact which led Evans-Pritchard to appeal to a residual "correspondence with reality" in distinguishing between "mystical" and "scientific" notions. The conception of reality is indeed indispensable to any understanding of the point of a way of life. But it is not a conception which can be explicated as Evans-Pritchard tries to explicate it, in terms of what science reveals to be the case; for a form of the conception of reality must already be presupposed before we can make any sense of the expression "what science reveals to be the case."

## II. OUR STANDARDS AND THEIRS

In Part I, I attempted, by analyzing a particular case, to criticize by implication a particular view of how we can understand a primitive institution. In this Part I shall have two aims. First, I shall examine in a more formal way a general philosophical argument, which attempts to show that

the approach I have been criticizing is in principle the right one. This argument has been advanced by Mr. Alasdair MacIntyre in two places: (a) in a paper entitled *Is Understanding Religion Compatible with Believing?* read to the Sesquicentennial Seminar of the Princeton Theological Seminar in 1962.<sup>22</sup> (b) In a contribution to *Philosophy, Politics and Society (Second Series)*,<sup>23</sup> entitled *A Mistake about Causality in Social Science*. Next, I shall make some slightly more positive suggestions about how to overcome the difficulty from which I started: how to make intelligible in our terms institutions belonging to a primitive culture, whose standards of rationality and intelligibility are apparently quite at odds with our own.

The relation between MacIntyre, Evans-Pritchard, and myself is a complicated one. MacIntyre takes Evans-Pritchard's later book, *Nuer Religion*, as an application of a point of view like mine in *The Idea of a Social Science*; he regards it as an object lesson in the absurd results to which such a position leads, when applied in practice. My own criticisms of Evans-Pritchard, on the other hand, have come from precisely the opposite direction. I have tried to show that Evans-Pritchard did not at the time of writing *The Azande* agree with me *enough*; that he did not take seriously enough the idea that the concepts used by primitive peoples can only be interpreted in the context of the way of life of those peoples. Thus I have in effect argued that Evans-Pritchard's account of the Azande is unsatisfactory precisely to the extent that he agrees with MacIntyre and not me.

The best point at which to start considering MacIntyre's position is that at which he agrees with me—in emphasizing the importance of possibilities of *description* for the concept of human action. An agent's action "is identified fundamentally as what it is by the description under which he deems it to fall." Since, further, descriptions must be intelligible to other people, an action "must fall under some description which is socially recognizable as the description of an action."<sup>24</sup> "To identify the limits of social action in a given period," therefore, "is to identify the stock of descriptions current in that age."<sup>25</sup> MacIntyre correctly points out that descriptions do not exist in isolation, but occur "as constituents of beliefs, speculations and projects."

<sup>21</sup> Notice that I have *not* said that Azande conceptions of witchcraft have nothing to do with understanding the world at all. The point is that a different form of the concept of understanding is involved here.

<sup>22</sup> To be published along with other papers, by the Macmillan Company.

<sup>23</sup> Edited by Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1962).

<sup>24</sup> *Ibid.*, p. 58.

<sup>25</sup> *Ibid.*, p. 60.

As these in turn "are continually criticized, modified, rejected, or improved, the stock of descriptions changes. The changes in human action are thus intimately linked to the thread of rational criticism in human history."

This notion of rational criticism, MacIntyre points out, requires the notion of choice between alternatives, to explain which "is a matter of making clear what the agent's criterion was and why he made use of this criterion rather than another and to explain why the use of this criterion appears rational to those who invoke it."<sup>26</sup> Hence "in explaining the rules and conventions to which action in a given social order conform (*sic*) we cannot omit reference to the rationality or otherwise of those rules and conventions." Further, "the beginning of an explanation of why certain criteria are taken to be rational in some societies is that they *are* rational. And since this has to enter into our explanation we cannot explain social behaviour independently of our own norms of rationality."

I turn now to criticism of this argument. Consider first MacIntyre's account of changes in an existing "stock" of available descriptions of actions. How does a candidate for inclusion *qualify* for admission to the stock? Unless there are limits, all MacIntyre's talk about possibilities of description circumscribing possibilities of action becomes nugatory, for there would be nothing to stop anybody inventing some arbitrary verbal expression, applying it to some arbitrary bodily movement, and thus adding that expression to the stock of available descriptions. But of course the new description must be an *intelligible* one. Certainly, its intelligibility cannot be decided by whether or not it belongs to an *existing* stock of descriptions, since this would rule out precisely what is being discussed: the addition of *new* descriptions to the stock. "What can intelligibly be said" is not equivalent to "what has been intelligibly said," or it would never be possible to say anything new. *Mutatis mutandis* it would never be possible to *do* anything new. Nevertheless the intelligibility of anything new said or done does depend in a certain way on what already has been said or done and understood. The crux of this problem lies in how we are to understand that "in a certain way."

In *Is Understanding Religion Compatible with Believing?* MacIntyre asserts that the development through criticism of the standards of intelligibility

current in a society is ruled out by my earlier account (in *The Idea of a Social Science*) of the origin in social institutions themselves of such standards. I shall not now repeat my earlier argument, but simply point out that I did, in various passages,<sup>27</sup> emphasize the *open* character of the "rules" which I spoke of in connection with social institutions: i.e., the fact that in changing social situations, reasoned decisions have to be made about what is to count as "going on in the same way." MacIntyre's failure to come to terms with this point creates difficulties for him precisely analogous to those which he mistakenly attributes to my account.

It is a corollary of his argument up to this point, as well as being intrinsically evident, that a new description of action must be intelligible to the members of the society in which it is introduced. On my view the point is that what determines this is the further development of rules and principles already implicit in the previous ways of acting and talking. To be emphasized are not the actual members of any "stock" of descriptions; but the *grammar* which they express. It is through this that we understand their structure and sense, their mutual relations, and the sense of new ways of talking and acting that may be introduced. These new ways of talking and acting may very well at the same time involve modifications in the grammar, but we can only speak thus if the new grammar is (to its users) intelligibly related to the old.

But what of the intelligibility of such changes to observers from another society with a different culture and different standards of intelligibility? MacIntyre urges that such observers must make clear "what the agent's criterion was and why he made use of this criterion rather than another and why the use of this criterion appears rational to those who invoke it." Since what is at issue is the precise relation between the concepts of rationality current in these different societies it is obviously of first importance to be clear about *whose* concept of rationality is being alluded to in this quotation. It seems that it must be that which is current in the society in which the criterion is invoked. Something can appear rational to someone only in terms of *his* understanding of what is and is not rational. If *our* concept of rationality is a different one from his, then it makes no sense to say that anything either does or does not appear rational to *him* in *our* sense.

When MacIntyre goes on to say that the observer

<sup>26</sup> *Ibid.*, p. 61.

<sup>27</sup> Pp. 57-65; 91-94; 121-123.

"cannot omit reference to the rationality or otherwise of those rules and conventions" followed by the alien agent, whose concept of rationality is now in question: ours or the agent's? Since the observer must be understood now as addressing himself to members of his own society, it seems that the reference must here be to the concept of rationality current in the observer's society. Thus there is a *non sequitur* in the movement from the first to the second of the passages just quoted.

MacIntyre's thought here and in what immediately follows, seems to be this. The explanation of why, in Society *S*, certain actions are taken to be rational, has got to be an explanation for us; so it must be in terms of concepts intelligible to us. If then, in the explanation, we say that in fact those criteria *are* rational, we must be using the word "*rational*" in *our* sense. For this explanation would require that we had previously carried out an independent investigation into the actual rationality or otherwise of those criteria, and we could do this only in terms of an understood concept of rationality—*our* understood concept of rationality. The explanation would run: members of Society *S* have seen to be the case something that we know to be the case. If "what is seen to be the case" is common to us and them, it must be referred to under the same concept for each of us.

But obviously this explanation is not open to us. For we start from the position that standards of rationality in different societies do not always coincide; from the possibility, therefore, that the standards of rationality current in *S* are different from our own. So we cannot assume that it will make sense to speak of members of *S* as discovering something which we have also discovered; such discovery presupposes initial conceptual agreement.

Part of the trouble lies in MacIntyre's use of the expression, "the rationality of criteria," which he does not explain. In the present context to speak thus is to cloak the real problem, since what we are concerned with are differences in *criteria of rationality*. MacIntyre seems to be saying that certain standards are taken as criteria of rationality because they *are* criteria of rationality. But whose?

There are similar confusions in MacIntyre's other paper: *Is Understanding Religion Compatible with Believing?* There he argues that when we detect an internal incoherence in the standards of intelligibility current in an alien society and try to show why this does not appear, or is made tolerable to that society's members, "we have already invoked our standards." In what sense is this true? Insofar

as we "detect" and "show" something, obviously we do so in a sense intelligible to us; so we are limited by what *counts* (for us) as "detecting," "showing" something. Further, it may well be that the interest in showing and detecting such things is peculiar to our society—that we are doing something in which members of the studied society exhibit no interest, because the institutions in which such an interest could develop are lacking. Perhaps too the pursuit of that interest in our society has led to the development of techniques of inquiry and modes of argument which again are not to be found in the life of the studied society. But it cannot be guaranteed in advance that the methods and techniques we have used in the past—e.g., in elucidating the logical structure of arguments in our own language and culture—are going to be equally fruitful in this new context. They will perhaps need to be extended and modified. No doubt, if they are to have a logical relation to our previous forms of investigation, the new techniques will have to be recognizably continuous with previously used ones. But they must also so extend our conception of intelligibility as to make it possible for us to see what intelligibility amounts to in the life of the society we are investigating.

The task MacIntyre says we must undertake is to make intelligible (a) (to us) why it is that members of *S* think that certain of their practices are intelligible (b) (to them), when in fact they are not. I have introduced differentiating letters into my two uses of "intelligible," to mark the complexity that MacIntyre's way of stating the position does not bring out: the fact that we are dealing with two different senses of the word "intelligible." The relation between these is precisely the question at issue. MacIntyre's task is not like that of making intelligible a natural phenomenon, where we are limited only by what counts as intelligibility for us. We must somehow bring *S*'s conception of intelligibility (b) into (intelligible!) relation with our own conception of intelligibility (a). That is, we have to create a new unity for the concept of intelligibility, having a certain relation to our old one and perhaps requiring a considerable realignment of our categories. We are not seeking a state in which things will appear to us just as they do to members of *S*, and perhaps such a state is unattainable anyway. But we *are* seeking a way of looking at things which goes beyond our previous way in that it has in some way taken account of and incorporated the other way that members of *S* have of looking at things. Seriously to study another way

of life is necessarily to seek to extend our own—not simply to bring the other way within the already existing boundaries of our own, because the point about the latter in their present form, is that they *ex hypothesi* exclude that other.

There is a dimension to the notions of rationality and intelligibility which may make it easier to grasp the possibility of such an extension. I do not think that MacIntyre takes sufficient account of this dimension and, indeed, the way he talks about “norms of rationality” obscures it. Rationality is not *just* a concept in a language like any other; it is this too, for, like any other concept it must be circumscribed by an established use: a use, that is, established in the language. But I think it is not a concept which a language may, as a matter of fact, have and equally well may not have, as is, for instance, the concept of politeness. It is a concept necessary to the existence of any language: to say of a society that it has a language<sup>28</sup> is also to say that it has a concept of rationality. There need not perhaps be any *word* functioning in its language as “rational” does in ours, but at least there must be features of its members’ use of language analogous to those features of *our* use of language which are connected with our use of the word “rational.” Where there is language it must make a difference what is said and this is only possible where the saying of one thing rules out, on pain of failure to communicate, the saying of something else. So in one sense MacIntyre is right in saying that we have already invoked our concept of rationality in saying of a collection of people that they constitute a society with a language: in the sense, namely, that we imply formal analogies between their behavior and that behavior in our society which we refer to in distinguishing between rationality and irrationality. This, however, is so far to say nothing about what in particular constitutes rational behavior in that society; that would require more particular knowledge about the norms they appeal to in living their lives. In other words, it is not so much a matter of invoking “our own norms of rationality” as of invoking our notion of rationality in speaking of their behavior in terms of “conformity to norms.” But how precisely this notion is to be applied to them will depend on our reading of their conformity to norms—what counts for them as conformity and what does not.

Earlier I criticized MacIntyre’s conception of a “stock of available descriptions.” Similar criticisms

apply to his talk about “our norms of rationality,” if these norms are taken as forming some finite set. Certainly we learn to think, speak, and act rationally *through* being trained to adhere to particular norms. But having learned to speak, etc., rationally does not *consist* in having been trained to follow those norms; to suppose that would be to overlook the importance of the phrase “and so on” in any description of what someone who follows norms does. We must, if you like, be open to new possibilities of what could be invoked and accepted under the rubric of “rationality”—possibilities which are perhaps suggested and limited by what we have hitherto so accepted, but not uniquely determined thereby.

This point can be applied to the possibilities of our grasping forms of rationality different from ours in an alien culture. First, as I have indicated, these possibilities are limited by certain formal requirements centering round the demand for consistency. But these formal requirements tell us nothing about what in particular is to *count* as consistency, just as the rules of the propositional calculus limit, but do not themselves determine, what are to be proper values of *p*, *q*, etc. We can only determine this by investigating the wider context of the life in which the activities in question are carried on. This investigation will take us beyond merely specifying the rules governing the carrying out of those activities. For, as MacIntyre quite rightly says, to note that certain rules are followed is so far to say nothing about the *point* of the rules; it is not even to decide whether or not they have a point at all.

MacIntyre’s recipe for deciding this is that “in bringing out this feature of the case one shows also whether the use of this concept is or is not a possible one for people who have the standards of intelligibility in speech and action which we have.”<sup>29</sup> It is important to notice that his argument, contrary to what he supposes, does not in fact show that our *own* standards of rationality occupy a peculiarly central position. The appearance to the contrary is an optical illusion engendered by the fact that MacIntyre’s case has been advanced in the English language and in the context of 20th Century European culture. But a formally similar argument could be advanced in *any* language containing concepts playing a similar role in that language to those of “intelligibility” and “rationality” in ours. This shows that, so far from over-

<sup>28</sup> I shall not discuss here what justifies us in saying *this* in the first place.

<sup>29</sup> *Is Understanding Religion Compatible with Believing?*

coming relativism, as he claims, MacIntyre himself falls into an extreme form of it. He disguises this from himself by committing the very error of which, wrongly as I have tried to show, he accuses me: the error of overlooking the fact that "criteria and concepts have a history." While he emphasizes this point when he is dealing with the concepts and criteria governing action in particular social contexts, he forgets it when he comes to talk of the *criticism* of such criteria. Do not the criteria appealed to in the criticism of existing institutions equally have a history? And in whose society do they have that history? MacIntyre's implicit answer is that it is in ours; but if we are to speak of difficulties and incoherencies appearing and being detected in the way certain practices have hitherto been carried on in a society, surely this can only be understood in connection with problems arising in the carrying on of the activity. Outside that context we could not begin to grasp what was problematical.

Let me return to the Azande and consider something which MacIntyre says about them, intended to support the position I am criticizing.

The Azande believe that the performance of certain rites in due form affects their common welfare; this belief cannot in fact be refuted. For they also believe that if the rites are ineffective it is because someone present at them had evil thoughts. Since this is always possible, there is never a year when it is unavoidable for them to admit that the rites were duly performed, but they did not thrive. Now the belief of the Azande is not unfalsifiable in principle (we know perfectly well what would falsify it—the conjunction of the rite, no evil thoughts and disasters). But in fact it cannot be falsified. Does this belief stand in need of rational criticism? And if so by what standards? It seems to me that one could only hold the belief of the Azande rational *in the absence* of any practice of science and technology in which criteria of effectiveness, ineffectiveness and kindred notions had been built up. But to say this is to recognize the appropriateness of scientific criteria of judgment from our standpoint. The Azande do not intend their belief either as a piece of science or as a piece of non-science. They do not possess these categories. It is only *post eventum*, in the light of later and more sophisticated understanding that their belief and concepts can be classified and evaluated at all.<sup>30</sup>

Now in one sense classification and evaluation of Zande beliefs and concepts does require "a more sophisticated understanding" than is found in Zande culture; for the sort of classification and

evaluation that are here in question are sophisticated philosophical activities. But this is not to say that Zande forms of life are to be classified and evaluated in the way MacIntyre asserts: in terms of certain specific forms of life to be found in our culture, according as they do or do not measure up to what is required within these. MacIntyre confuses the sophistication of the interest in classification with the sophistication of the concepts employed in our classificatory work. It is of interest to us to understand how Zande magic is related to science; the concept of such a comparison is a very sophisticated one; but this does not mean that we have to see the unsophisticated Zande practice in the light of more sophisticated practices in our own culture, like science—as perhaps a more primitive form of it. MacIntyre criticizes, justly, Sir James Frazer for having imposed the image of his own culture on more primitive ones; but that is exactly what MacIntyre himself is doing here. It is extremely difficult for a sophisticated member of a sophisticated society to grasp a very simple and primitive form of life: in a way he must jettison his sophistication, a process which is itself perhaps the ultimate in sophistication. Or, rather, the distinction between sophistication and simplicity becomes unhelpful at this point.

It may be true, as MacIntyre says, that the Azande do not have the categories of science and non-science. But Evans-Pritchard's account shows that they do have a fairly clear working distinction between the technical and the magical. It is neither here nor there that individual Azande may sometimes confuse the categories, for such confusions may take place in any culture. A much more important fact to emphasize is that *we* do not initially have a category that looks at all like the Zande category of magic. Since it is we who want to understand the Zande category, it appears that the onus is on us to extend our understanding so as to make room for the Zande category, rather than to insist on seeing it in terms of our own ready-made distinction between science and non-science. Certainly the sort of understanding we seek requires that we see the Zande category in relation to our own already understood categories. But this neither means that it is right to "evaluate" magic in terms of criteria belonging to those other categories; nor does it give any clue as to *which* of our existing categories of thought will provide the best point of reference from which we can understand the point of the Zande practices.

<sup>30</sup> *Ibid.*

MacIntyre has no difficulty in showing that if the rites which the Azande perform in connection with their harvests are "classified and evaluated" by reference to the criteria and standards of science or technology, then they are subject to serious criticism. He thinks that the Zande "belief" is a sort of *hypothesis* like, e.g., an Englishman's belief that all the heavy rain we have been having is due to atomic explosions.<sup>31</sup> MacIntyre believes that he is applying as it were a neutral concept of "*A* affecting *B*," equally applicable to Zande magic and western science. In fact, however, he is applying the concept with which *he* is familiar, one which draws its significance from its use in scientific and technological contexts. There is no reason to suppose that the Zande magical concept of "*A* affecting *B*" has anything like the same significance. On the contrary, since the Azande do, in the course of their practical affairs, apply something very like our technical concept—though perhaps in a more primitive form—and since their attitude to and thought about their magical rites are quite different from those concerning their technological measures, there is every reason to think that their concept of magical "influence" is quite different. This may be easier to accept if it is remembered that, even in our own culture, the concept of causal influence is by no means monolithic: when we speak, for example, of "what made Jones get married," we are not saying the same kind of thing as when we speak of "what made the aeroplane crash"; I do not mean simply that the events of which we speak are different in kind but that the relation between the events is different also. It should not then be difficult to accept that in a society with quite different institutions and ways of life from our own, there may be concepts of "causal influence" which behave even more differently.

But I do not want to say that we are quite powerless to find ways of thinking in our own society that will help us to see the Zande institution in a clearer light. I only think that the direction in which we should look is quite different from what MacIntyre suggests. Clearly the nature of Zande life is such that it is of very great importance to them that their crops should thrive. Clearly too

they take all kinds of practical "technological" steps, within their capabilities, to ensure that they *do* thrive. But that is no reason to see their magical rites as a further, misguided such step. A man's sense of the importance of something to him shows itself in all sorts of ways: not merely in precautions to safeguard that thing. He may want to come to terms with its importance to him in quite a different way: to contemplate it, to gain some sense of his life in relation to it. He may wish thereby, in a certain sense, to *free* himself from dependence on it. I do not mean by making sure that it does not let him down, because the point is that, *whatever* he does, he may still be let down. The important thing is that he should understand *that* and come to terms with it. Of course, merely to understand that is not to come to terms with it, though perhaps it is a necessary condition for so doing, for a man may equally well be transfixed and terrorized by the contemplation of such a possibility. He must see that he can still go on even if he is let down by what is vitally important to him; and he must so order his life that he still *can* go on in such circumstances. I stress once again that I do not mean this in the sense of becoming "technologically independent," because from the present point of view technological independence is yet another form of dependence. Technology destroys some dependencies but always creates new ones, which may be fiercer—because harder to understand—than the old. This should be particularly apparent to us.<sup>32</sup>

In Judaeo-Christian cultures the conception of "If it be Thy Will," as developed in the story of Job, is clearly central to the matter I am discussing. Because this conception is central to Christian prayers of supplication, they may be regarded from one point of view as freeing the believer from dependence on what he is supplicating for.<sup>33</sup> Prayers cannot play this role if they are regarded as a means of influencing the outcome for in that case the one who prays is still dependent on the outcome. He frees himself from this by acknowledging his complete dependence on God; and this is totally unlike any dependence on the outcome precisely because God is eternal and the outcome contingent.

I do not say that Zande magical rites are at all like Christian prayers of supplication in the positive

<sup>31</sup> In what follows I have been helped indirectly, but greatly, by some unpublished notes made by Wittgenstein on Frazer, which Mr. Rush Rhees was kind enough to show me; and also by various scattered remarks on folklore in *The Notebooks of Simone Weil* (London, Routledge & Kegan Paul, 1963).

<sup>32</sup> The point is beautifully developed by Simone Weil in her essay on "The Analysis of Oppression" in *Oppression and Liberty* (London, Routledge & Kegan Paul, 1958).

<sup>33</sup> I have been helped to see this point by a hitherto unpublished essay on the concept of prayer by Mr. D. Z. Phillips.

attitude to contingencies which they express. What I do suggest is that they are alike in that they do, or may, express an attitude to contingencies; one, that is, which involves recognition that one's life is subject to contingencies, rather than an attempt to control these. To characterize this attitude more specifically one should note how Zande rites emphasize the importance of certain fundamental features of their life which MacIntyre ignores. MacIntyre concentrates implicitly on the relation of the rites to consumption, but of course they are also fundamental to social relations and this seems to be emphasized in Zande notions of witchcraft. We have a drama of resentments, evil-doing, revenge, expiation, in which there are ways of dealing (symbolically) with misfortunes and their disruptive effect on a man's relations with his fellows, with ways in which life can go on despite such disruptions.

How is my treatment of this example related to the general criticisms I was making of MacIntyre's account of what it is for us to see the *point* of the rules and conventions followed in an alien form of life? MacIntyre speaks as though our own rules and conventions are somehow a paradigm of what it is for rules and conventions to have a point, so that the only problem that arises is in accounting for the point of the rules and conventions in some other society. But in fact, of course, the problem is the same in relation to our own society as it is in relation to any other; no more than anyone else's are *our* rules and conventions immune from the danger of being or becoming pointless. So an account of this matter cannot be given simply in terms of any set of rules and conventions at all: our own or anyone else's; it requires us to consider the relation of a set of rules and conventions to something else. In my discussion of Zande magical rites just now what I tried to relate the magical rites to was a sense of the significance of human life. This notion is, I think, indispensable to any account of what is involved in understanding and learning from an alien culture; I must now try to say more about it.

In a discussion of Wittgenstein's philosophical use of language games<sup>34</sup> Mr. Rush Rhees points out that to try to account for the meaningfulness of language solely in terms of isolated language games is to omit the important fact that ways of speaking are not insulated from each other in mutually exclusive systems of rules. What can be said in one context by the use of a certain expression depends for its sense on the uses of that

expression in other contexts (different language games). Language games are played by men who have lives to live—lives involving a wide variety of different interests, which have all kinds of different bearings on each other. Because of this, what a man says or does may make a difference not merely to the performance of the activity upon which he is at present engaged, but to his *life* and to the lives of other people. Whether a man sees point in what he is doing will then depend on whether he is able to see any unity in his multifarious interests, activities, and relations with other men; what sort of sense he sees in his life will depend on the nature of this unity. The ability to see this sort of sense in life depends not merely on the individual concerned, though this is not to say it does not depend on him at all; it depends also on the possibilities for making such sense which the culture in which he lives does, or does not, provide.

What we may learn by studying other cultures are not merely possibilities of different ways of doing things, other techniques. More importantly we may learn different possibilities of making sense of human life, different ideas about the possible importance that the carrying out of certain activities may take on for a man, trying to contemplate the sense of his life as a whole. This dimension of the matter is precisely what MacIntyre misses in his treatment of Zande magic: he can see in it only a (misguided) technique for producing consumer goods. But a Zande's crops are not just potential objects of consumption: the life he lives, his relations with his fellows, his chances for acting decently or doing evil, may all spring from his relation to his crops. Magical rites constitute a form of expression in which these possibilities and dangers may be contemplated and reflected on—and perhaps also thereby transformed and deepened. The difficulty we find in understanding this is not merely its remoteness from science, but an aspect of the general difficulty we find, illustrated by MacIntyre's procedure, of thinking about such matters at all except in terms of "efficiency of production"—production, that is, for consumption. This again is a symptom of what Marx called the "alienation" characteristic of man in industrial society, though Marx's own confusions about the relations between production and consumption are further symptoms of that same alienation. Our blindness to the point of primitive modes of life is a corollary of the pointlessness of much of our own life.

<sup>34</sup> Rush Rhees, "Wittgenstein's Builders," *Proceedings of the Aristotelian Society*, vol. 20 (1960), pp. 171-186.

I have now explicitly linked my discussion of the "point" of a system of conventions with conceptions of good and evil. My aim is not to engage in moralizing, but to suggest that the concept of *learning from* which is involved in the study of other cultures is closely linked with the concept of *wisdom*. We are confronted not just with different techniques, but with new possibilities of good and evil, in relation to which men may come to terms with life. An investigation into this dimension of a society may indeed require a quite detailed inquiry into alternative techniques (e.g., of production), but an inquiry conducted for the light it throws on those possibilities of good and evil. A very good example of the kind of thing I mean is Simone Weil's analysis of the techniques of modern factory production in *Oppression and Liberty*, which is not a contribution to business management, but part of an inquiry into the peculiar form which the evil of oppression takes in our culture.

In saying this, however, I may seem merely to have lifted to a new level the difficulty raised by MacIntyre of how to relate our own conceptions of rationality to those of other societies. Here the difficulty concerns the relation between our own conceptions of good and evil and those of other societies. A full investigation would thus require a discussion of ethical relativism at this point. I have tried to show some of the limitations of relativism in an earlier paper.<sup>36</sup> I shall close the present essay with some remarks which are supplementary to that.

I wish to point out that the very conception of human life involves certain fundamental notions—which I shall call "limiting notions"—which have an obvious ethical dimension, and which indeed in a sense determine the "ethical space," within which the possibilities of good and evil in human life can be exercised. The notions which I shall discuss very briefly here correspond closely to those which Vico made the foundation of his idea of natural law, on which he thought the possibility of understanding human history rested: birth, death, sexual relations. Their significance here is that they are inescapably involved in the life of all known human societies in a way which gives us a clue where to look, if we are puzzled about the point of an alien system of institutions. The specific forms which these concepts take, the particular institutions in which they are expressed, vary very considerably from one society to another; but their central position within a society's institutions is and must be a constant

factor. In trying to understand the life of an alien society, then, it will be of the utmost importance to be clear about the way in which these notions enter into it. The actual practice of social anthropologists bears this out, although I do not know how many of them would attach the same kind of importance to them as I do.

I speak of a "limit" here because these notions, along no doubt with others, give shape to what we understand by "human life"; and because a concern with questions posed in terms of them seems to me constitutive of what we understand by the "morality" of a society. In saying this, I am of course disagreeing with those moral philosophers who have made attitudes of approval and disapproval, or something similar, fundamental in ethics, and who have held that the *objects* of such attitudes were conceptually irrelevant to the conception of morality. On that view, there might be a society where the sorts of attitude taken up in *our* society to questions about relations between the sexes were reserved, say, for questions about the length people wear their hair, and *vice versa*. This seems to me incoherent. In the first place, there would be a confusion in *calling* a concern of that sort a "moral" concern, however passionately felt. The story of Samson in the Old Testament confirms rather than refutes this point, for the interdiction on the cutting of Samson's hair is, of course, connected there with much else: and pre-eminently, it should be noted, with questions about sexual relations. But secondly, if that is thought to be merely verbal quibbling, I will say that it does not seem to me a merely conventional matter that T. S. Eliot's trinity of "birth, copulation and death" happen to be such deep objects of human concern. I do not mean just that they are made such by fundamental psychological and sociological forces, though that is no doubt true. But I want to say further that the very notion of human life is limited by these conceptions.

Unlike beasts, men do not merely live but also have a conception of life. This is not something that is simply added to their life; rather, it changes the very sense which the word "life" has, when applied to men. It is no longer equivalent to "animate existence." When we are speaking of the life of man, we can ask questions about what is the right way to live, what things are most important in life, whether life has any significance, and if so what.

To have a conception of life is also to have a

<sup>36</sup> Peter Winch, "Nature and Convention," *Proceedings of the Aristotelian Society*, vol. 20 (1960), pp. 231-252.

conception of death. But just as the "life" that is here in question is not the same as animate existence, so the "death" that is here in question is not the same as the end of animate existence. My conception of the death of an animal is of an event that will take place in the world; perhaps I shall observe it—and my life will go on. But when I speak of "my death," I am not speaking of a future event in my life;<sup>36</sup> I am not even speaking of an event in anyone else's life. I am speaking of the cessation of my world. That is also a cessation of my ability to do good or evil. It is not just that *as a matter of fact* I shall no longer be able to do good or evil after I am dead; the point is that my very *concept* of what it is to be able to do good or evil is deeply bound up with my concept of my life as ending in death. If ethics is a concern with the right way to live, then clearly the nature of this concern must be deeply affected by the concept of life as ending in death. One's attitude to one's life is at the same time an attitude to one's death.

This point is very well illustrated in an anthropological datum which MacIntyre confesses himself unable to make any sense of.

According to Spencer and Gillen some aborigines carry about a stick or stone which is treated *as if* it is or embodies the soul of the individual who carries it. If the stick or stone is lost, the individual anoints himself as the dead are anointed. Does the concept of "carrying one's soul about with one" make sense? Of course we can redescribe what the aborigines are doing and transform it into sense, and perhaps Spencer and Gillen (and Durkheim who follows them) misdescribe what occurs. But if their reports are not erroneous, we confront a blank wall here, so far as meaning is concerned, although it is easy to give the rules for the use of the concept.<sup>37</sup>

MacIntyre does not say why he regards the concept of carrying one's soul about with one in a stick "thoroughly incoherent." He is presumably influenced by the fact that it would be hard to make sense of an action like this if performed by a twentieth-century Englishman or American; and by the fact that the soul is not a material object like a piece of paper and cannot, therefore, be carried about in a stick as a piece of paper might be. But it does not seem to me so hard to see sense in the practice, even from the little we are told

about it here. Consider that a lover in our society may carry about a picture or lock of hair of the beloved; that this may symbolize for him his relation to the beloved and may, indeed, change the relation in all sorts of ways: for example, strengthening it or perverting it. Suppose that when the lover loses the locket he feels guilty and asks his beloved for her forgiveness: there might be a parallel here to the aboriginal's practice of anointing himself when he "loses his soul." And is there necessarily anything irrational about either of these practices? Why should the lover not regard his carelessness in losing the locket as a sort of betrayal of the beloved? Remember how husbands and wives may feel about the loss of a wedding ring. The aborigine is clearly expressing a concern with his life as a whole in this practice; the anointing shows the close connection between such a concern and contemplation of death. Perhaps it is precisely this practice which makes such a concern possible for him, as religious sacraments make certain sorts of concern possible. The point is that a concern with one's life as a whole, involving as it does the limiting conception of one's death, if it is to be expressed *within* a person's life, can necessarily only be expressed quasi-sacramentally. The form of the concern shows itself in the form of the sacrament.

The sense in which I spoke also of sex as a "limiting concept" again has to do with the concept of a human life. The life of a man is a man's life and the life of a woman is a woman's life: the masculinity or the femininity are not just *components* in the life, they are its *mode*. Adapting Wittgenstein's remark about death, I might say that my masculinity is not an experience in the world, but my way of experiencing the world. Now the concepts of masculinity and femininity obviously require each other. A man is a man in relation to women; and a woman is a woman in relation to men.<sup>38</sup> Thus the form taken by man's relation to women is of quite fundamental importance for the significance he can attach to his own life. The vulgar identification of morality with sexual morality certainly is vulgar; but it is a vulgarization of an important truth.

The limiting character of the concept of birth is obviously related to the points I have sketched regarding death and sex. On the one hand, my

<sup>36</sup> Cf. Wittgenstein, *Tractatus Logico-Philosophicus*, 6.431–6.4311.

<sup>37</sup> *Is Understanding Religion Compatible with Believing?*

<sup>38</sup> These relations, however, are not simple converses. See Georg Simmel, "Das Relative und das Absolute im Geschlechter-Problem" in *Philosophische Kultur* (Leipzig, Werner Klinkhardt, 1911).

birth is no more an event in my life than is my death; and through my birth ethical limits are set for my life quite independently of my will: I am, from the outset, in specific relations to other people, from which obligations spring which cannot but be ethically fundamental.<sup>39</sup> On the other hand, the concept of birth is fundamentally linked to that of relations between the sexes. This remains true, however much or little may be known in a society about the contribution of males and females to procreation; for it remains true that man is born of woman, not of man. This, then, adds a new dimension to the ethical institutions in which relations between the sexes are expressed.

I have tried to do no more, in these last brief remarks, than to focus attention in a certain direction. I have wanted to indicate that forms of these limiting concepts will necessarily be an important feature of any human society and that conceptions of good and evil in human life will necessarily be connected with such concepts. In any attempt to understand the life of another society, therefore, an investigation of the forms taken by such concepts—their role in the life of the society—must always take a central place and

provide a basis on which understanding may be built.

Now since the world of nations has been made by men, let us see in what institutions men agree and always have agreed. For these institutions will be able to give us the universal and eternal principles (such as every science must have) on which all nations were founded and still preserve themselves.

We observe that all nations, barbarous as well as civilized, though separately founded because remote from each other in time and space, keep these three human customs: all have some religion, all contract solemn marriages, all bury their dead. And in no nation, however savage and crude, are any human actions performed with more elaborate ceremonies and more sacred solemnity than the rites of religion, marriage and burial. For by the axiom that "uniform ideas, born among peoples unknown to each other, must have a common ground of truth," it must have been dictated to all nations that from these institutions humanity began among them all, and therefore they must be most devoutly guarded by them all, so that the world should not again become a bestial wilderness. For this reason we have taken these three eternal and universal customs as the first principles of this Science.<sup>40</sup>

*University College of Swansea,  
University of Wales*

<sup>39</sup> For this reason, among others, I think A. I. Melden is wrong to say that parent-child obligations and rights have nothing directly to do with physical genealogy. Cf. Melden, *Rights and Right Conduct* (Oxford, Basil Blackwell, 1959).

<sup>40</sup> Giambattista Vico, *The New Science*, §§ 332–333.

## IV. TOWARD A PHENOMENOLOGY OF EXPERIENCE

HERBERT SPIEGELBERG

THIS paper\* does not aim at a comprehensive phenomenology of experience. Its main purpose is to show how the phenomenological approach can add to a fuller understanding of a phenomenon whose prestige stands almost in inverse proportion to its lucidity. "Experience" is easily the most eulogized title of contemporary philosophy. Hence it has become its most over-extended and diluted concept. Any attempt to clarify the nature of the phenomenon intensively and to examine the rights of experience must therefore begin with a demarcation of its field.

### I. THE PROBLEM OF DENOTATION

I shall attempt this demarcation of the denotational field of the term "experience" by telling what I intend to exclude from it. Such a procedure may appear at first highhanded and only of private interest. However, the lines I shall draw are based on previous examination of "family resemblances" which make the incisions in the continuum of the phenomena a matter not only of convenience but of requiredness: they indicate the places where the articulation of the phenomena calls for such incisions.

I am not going to consider here with John Dewey "all the complex series of transactions which occur between the live creature and its environment."<sup>1</sup> Such a definition, at least if taken literally, would include mere physico-chemical exchanges of energy between the organism and its surroundings, e.g., a plant absorbing carbon dioxide and discharging oxygen. I shall disregard also "experiences" such as accidents of good or bad luck as long as they remain unnoticed. And I shall finally ignore "experience" in the sense of the stock of more or less wobbly generalizations, maxims, or "knack" which may make a person "experienced." No doubt there are connections between all these uses.

But for the purposes of the present discussion I want to focus on what, if I am not mistaken, the classical debates about experience have always been about: the kind of individual occurrences in which a potential knower makes cognitive contact with an individualized object. However, I want to make it clear at once that this characterization is not meant as a definition of the essence of experience but merely as an attempt to stake out the field for structural investigation, which eventually may lead to a structural definition. I shall use as a paradigm of such an experience the reader's awareness of the page on which these sentences are printed with its black marks against a white ground, beginning with the visual experience, matched by the experience of absorbing its meaning and by whatever other experiences may follow upon this reading experience.

### II. THE RECORD OF PHENOMENOLOGY

I shall start out with a brief account of the phenomenological exploration of experience thus far. Such an account seems to be particularly indicated since phenomenology is often suspected of being hostile to experience.

It is, therefore, worth stressing that Brentano, the major inspirer of Husserl's phenomenology, started his reform of philosophy with a new psychology "on empirical foundations," taking experience, to be sure, in a sense which included certain *a priori* insights. Husserl, in setting forth the idea of his pure phenomenology coupled this with an attack on empiricism, an attack intent on safeguarding the right of insights into general essences against the monopoly of experience of merely individual facts. It is all the more important to realize that in spite of this anti-empiricist front Husserl subscribed to the positivist principle of unconditional adherence to the given without addition or diminution, so much so that he affirmed: "we

\* This article is a development of a paper which was read in a symposium on Experience before the Metaphysical Society of America at Notre Dame University in March 1960.

<sup>1</sup> See, e.g., Gail Kennedy, "Dewey's Concept of Experience: Determinate, Indeterminate, and Problematic," *Journal of Philosophy*, vol. 56 (1959), p. 802.

phenomenologists are the true positivists."<sup>2</sup> But the decisive evidence is his actual approach to experience in his phenomenology of consciousness. For here the study of the essential structure of experience is one of Husserl's major concerns, and this, among other things, for the weighty reason that perceptual experience is the only valid evidence for our knowledge of the "real" world. Husserl's studies on *Experience and Judgment*, as developed under his supervision by Ludwig Landgrebe and published posthumously,<sup>3</sup> contain in fact some of his most concrete pieces of phenomenologizing, precisely in their sections on experience. Specifically, they show pre-predicative experience of our life-world as the matrix of all our propositional and scientific knowledge.

The same interest in a widened and deepened conception of "phenomenological experience" is manifested in Max Scheler's writings, particularly in those dealing with our experience of value. More recently, among the French phenomenological existentialists, Maurice Merleau-Ponty devoted much of his attention to this subject. His approach could perhaps be characterized best by a formula like "Against Empiricism; for lived experience (*expérience vécue*)," i.e., for experience as the manner in which each existing perceiver lives his body and his world.

What I am going to present in the following sections does not pretend to be a faithful exposition of common phenomenological doctrine or of that of any one of these phenomenologists. However, I have made a special effort to assimilate into my version what I consider the most valid parts of Husserl's pioneering insights. These seem to me often unduly neglected even by phenomenologists. But I must warn that whatever strange novelties I shall seem to peddle may turn out to be heresies which the master could rightfully disclaim. He should, however, receive the primary credit for whatever may appeal to the reader.

### III. SOME PHENOMENOLOGICAL FINDINGS ABOUT EXPERIENCE

An attempt to present merely what is distinctive about the phenomenological conception of experience would presuppose a panoramic acquaintance with all the classical and recent literature in the field. This, I am afraid, is beyond the scope of this paper. But, trying to make a virtue of an impossi-

bility, I might actually do better by simply presenting what seem to phenomenologists some of the central characteristics of experience, leaving it to the reader to decide whether there is anything original and possibly helpful about them. I shall, however, make an underhanded attempt to stress features which I suspect will be of particular interest and possible nuisance value to professed empiricists. And I shall do so by formulating eight compact theses on which I intend to comment as I go along.

#### 3.1. *Experience is an "intentional" act in which an experiencer is directed toward an experienced object.*

No different from other intentional acts, such as imagination, thought, or practical "intentions," experience refers to objects other than itself of which it is experience. In this sense of a reference beyond itself, experience has essentially a minimum of meaning, and meaningless experience is a contradiction in terms. The articulation of the global phenomenon in which experience occurs into the intending experience and the intended object of experience involves that any phenomenological study of experience has to pay parallel and equal attention to what is experienced just as it is experienced and to the ways in which our experiencing approaches it. Thus in describing the experience of this page we shall have to attend not only to the phenomenal characteristics, the perspective aspects and the modes of clarity with which it is given but equally to our ways of looking at it from various angles successively and in different attitudes, doing it, for instance, searchingly or doing it merely casually.

It should be realized that common parlance does not restrict the use of the word "experience" to acts of experiencing but that very often we also call the object or "content" of this act "an experience." Nevertheless, it is only by the grace of the experiencing act that the experienced object receives its title of being "an experience." Even in epistemological discussions of experience we often understand the word in a more global sense in which act and "content" are embraced by the term. It is all the more important not to overlook the articulation within this global phenomenon and not to lump them together. More important, we must be aware that our experiencing is set off against the object of our experience in a peculiar and unique manner. Neither is part of the other.

<sup>2</sup> *Ideen zu einer reinen Phänomenologie*, § 20.

<sup>3</sup> *Erfahrung und Urteil* (Prague, Academia, 1939).

3.2. *A full experience is a synthesis of several intentional acts.*

Our experiences do not simply string along in single file. They are object-focused. And they are not always directed toward separate objects. Often, and perhaps in most cases, they have identical foci. In other words, the intentions of several experiencing acts may converge upon one and the same object of experience. It is only when these acts coalesce in this manner that we can properly speak of having experience of an object. The experience of the present page implies a whole series of intentional acts, whose concentric meaning has been formed by a synthesis of all the impressions the reader has had since he first "laid eyes on it," moved toward it, or turned it.

3.3. *The experience of an object refers beyond itself.*

No experience is sharply circumscribed by precise contours. At its periphery each content melts into a background of dimly given data, which in turn refer beyond themselves toward other potential data; and each act of experiencing verges on other acts which may present more and different aspects of the same object. Each phase of an experience is thus surrounded by tracks or pathways to only partially given and even to not given aspects. Such references can then be fulfilled or disappointed in subsequent phases of experience. In the reading experience of this page, where our moving glance focuses on a limited area of the text, we find ourselves referred backward and forward over the page to what is thus far only marginally given and beyond to what is still completely beyond our glance. We are also referred to the reverse of the page, to preceding and subsequent pages, and to the whole magazine where this text appears.

3.4. *Experience has no temporal structure.*

There is no such thing as a merely instantaneous experience. Rather, experience has a temporal pattern opening from the very start toward future phases and subsequently also toward past phases. These past phases drop off constantly toward deeper levels of consciousness, where they are submerged by more recent arrivals, and become "sedimented," to use Husserl's picturesque expression. They are at first still within the range of immediate retention, but soon they are available

only to the recuperating effort of recollection. Future phases stand before us only in the mode of approaching expectation. The experience of this text at the present moment is suffused by at least several minutes of the reader's first and second impressions, and "irradiated" by what he may still experience if he feels sufficiently encouraged and at leisure to read on.

3.5. *Experience extends to any type of individual objects.*

Experience is not restricted to so-called sense-experience.<sup>4</sup> There is for the phenomenologist experience of relations, meanings, values, requiredness, other minds, social and cultural phenomena. Any kind of cognitive contact with particular data is an occasion for genuine experience. We experience not only each other's epidermises (or rather, their perspective appearances) but also each other's global bodies, personalities, and maybe even thoughts and feelings.

3.6. *Experience forms the pre-predicative stage of our cognitive life.*

The predicative stage of judgments and propositions with its polarization into subject and predicate differs essentially from the unpolarized structure of our immediate experience. But predicative knowledge has its primary foundation in such direct experience. Perception is its primary form, but so are feeling and other first-hand ("primordial") acts. The way to judgmental knowledge is prepared by a sequence of articulating acts of which Husserl has given some highly original accounts. From the plain seizure of the originally given via its retentive inspection and explication for its detailed structure we gradually approach the stage of "substratum" and "determination" which allows for expression in logical predication.

Illustrating at least something of these highly complex relationships I shall simply call attention to the situation in which we find ourselves when, after having spent some time in a room, we want to describe it. Evidently this presupposes an explicit refocusing and reorganization of the first experience.

3.7. *Experience constitutes the experienced.*

In introducing Husserl's term "constitution" I have to point out immediately that it allows for a number of more or less extreme interpretations,

<sup>4</sup> "So-called"—for what entitles us to call it *sense-experience*? What do we know about our senses that accounts for the structure and the content of these experiences? And how do we know about these "senses"?

and that it has given rise to considerable confusion and controversy among his followers. I shall here confine myself to a minimum sense of the term under which it means no more than that each object of experience establishes itself, or "settles" in our experience by taking shape before our eyes, as it were. Again the example of this magazine should help to make this clearer. Before encountering it we had a number of more or less vague anticipations based on such clues as our previous acquaintance with magazines and more specific data, which provided something like the steel frame of a skyscraper. But when we got sight of it first we had little more of a picture than that of a block of paper. As we opened it, this block rapidly took on more definite characteristics. Empty lots in the perceptual field were filled with pages, headlines, text, etc. At the present stage at least a few perspectives of this page have established themselves for the reader, sufficient to picture how the page would look from the bottom, etc. But even then many blanks remain to be filled before the picture of the entire article will stand complete before the reader's mind.

3.8. *Experience is a combination of receptive and spontaneous processes.*

While there are certainly receptive phases in all experience, especially in its initial stages, it would be quite mistaken to see in experience nothing but a passive process. Experience is permeated by spontaneous activities which Husserl called achievements (*Leistungen*). Kant's transcendental analysis of experience opened up the field for their investigation. Husserl, in tracing the active achievements of *Erfahrung*, described specifically the exploratory inspection, the laying hold and keeping hold of the object of experience, and its "explication." To him there is an entire hierarchy of constitutive acts conditioned by the first occurrence of empirical raw material. In sum, experience is an intertwined network of receptive and spontaneous processes, of undergoing and doing, in which the active phases are in the ascendancy the closer experience approaches the predicative stage of articulated knowledge.

To appreciate this we hardly need remind the reader of the way in which he first experienced this page and still continues doing so. After receiving his first impressions from it, he let his eyes range adventurously over its expanse before he began, more or less methodically, to look it over in an orderly fashion.

#### IV. THE GIVEN AND THE FOUND IN EXPERIENCE

Closer attention to the active factors in experience suggests a reconsideration of the concept of "the given." Among the many things taken for granted in our thinking about experience is the belief that the object of experience is "given." At least some philosophers, mostly critics of phenomenology, realize that the term "given" is highly ambiguous. Thus it may stand either for the sense data (whatever they may be), for the perspective appearances of an object, for the sides which they present, or for the global object with all its sides. But few if any students of experience seem to have inquired what the very idea of givenness implies. Obviously the term is a metaphor. How many of the features of the idea of "giving" or donating genuinely apply to the given as it occurs in experience?

Now the metaphor of giving or donating implies first of all that the initiative for the process lies with the giver, that the gift passes from him to the receiver or addressee. His role in such a process may seem at first to be primarily passive. Yet actually there can be no gift without some activity on the part of the recipient. Not only before the law, a gift cannot be accomplished without being accepted; without acceptance the gift fails, drops to the floor, as it were. One may try to force a gift on a potential receiver; but it is obvious that something forced down his throat is no longer a gift. There can be no *giving* without free *receiving*. Otherwise the given is at best the offered. But what exactly is accepting? It presupposes a certain openness and readiness on the part of the receiver to take over and hold on to what has been handed to him. This is certainly more than a merely passive disposition. Then, the recipient has to actually "open his hands," "close them" around the gift, and to "keep them closed" until he can deposit it within his domain. Other gestures are of course equally expressive of the act so symbolized.

How far do these features of the metaphor apply to the act of experience? In important regards they do. Whenever a new datum enters our ken, it is not the experiencer who has the initiative. He is certainly passive to the extent that he undergoes the impact of the "first impression," and that he continues to be affected by the influx of subsequent information. At this stage, it is the experienced object which does all the talking. On the other hand, it is very definitely true of experience that nothing can be given to the experiencer without his

openness, his readiness, and his willingness to receive. To the non-receptive subject, to the merely captive audience, nothing can be given, it can only be offered to it, and then it is mostly wasted. Now openness and readiness is to a considerable extent a matter of active control: we can open (or close) our mind and we can get set for an experience (and just as well guard ourselves against an experience). In addition to this phase, experiencing always requires the active operation in which we lay hold of or assimilate, whatever enters our opened field of experience.

But, is this kind of receptive experience, an experience mainly of the wait-and-see type, the only type of experience to be considered? Actually the very word "experience" and its equivalents in other languages suggest something much more active. For "ex-perience" in English refers to the extract or outcome of a special trial (the stem of the word is related to the Greek verb *peirōō*): hence the literal meaning of the word comes very close to experiment. Similarly in German the word "*erfahren*" includes the root "*fahren*," i.e., to drive, to travel. Hence *erfahren* expresses originally the attainment of knowledge by an outgoing effort.

Now there is indeed a type of experience in which the primary initiative lies with the experiencer. Such experience is based on seeking, not merely waiting, and terminates in finding, not merely receiving. This "finding experience" deserves separate and detailed exploration, since this is clearly the type which is at the base of all research experience. To this extent the found is even epistemologically more important than the given.<sup>5</sup> In the present context I can give only a first sketch of this type of experience.

Can there be such a thing as seeking without an object sought? Plato's *Meno* denies this, thus preparing the ground for his doctrine of innate ideas. I submit that there are at least two basic types of seeking. One is the merely exploratory, curiosity-directed surveying of new ground without looking for anything in particular. The other is definitely a search "for," directed by some more or less definite pattern of what we are looking for. Such a search and only such a search can be successful and unsuccessful. There are of course all

forms of transition. At times we may not even know what we are looking for until we find out by a discovery "This is what we had been looking for." In fact we cannot even be happily or unhappily surprised in our search if we have not at least "the remotest idea" of what we are after. But there is the limiting case of the merely mustering search in simple surveying, free of expectations, that can be fulfilled or disappointed. This, however, does not eliminate the possibility of the kind of surprises which result from a sudden change in the pattern of experiences to which we have become accustomed.

Now what is found as a result of a merely exploratory search constitutes itself on the basis of an approach which simply follows the lead of the phenomena. To be sure, finding something in the uncharted sea around us implies some criterion of what could be considered a find, as distinguished from our experience in those intervals of our exploration where we find "nothing." Somehow the found must present itself as sufficiently different, set off against the ground of the preceding experiences to give us the idea that here is something new. Then we can match our findings with each phase of the probe and chart the results.

The situation differs considerably when the search is undertaken under the guidance of a pattern of what we are seeking for. Not only shall we be looking for clues in the phenomena which will fit into this pattern, we shall also be exploring the field in a very different manner, following up leads in accordance with the pattern to which we want to match our findings. The found itself will present itself from the start with a fringe of references in anticipation of what is to be found next and what may fulfill or disappoint our expectations. True, such a pattern of anticipation may very well prejudice our search. But the found is still master of itself, though it is not *our* master. We are no longer at its mercy but can put it on the witness stand, as it were.

## V. TOWARD A PHENOMENOLOGY OF THE CONTEXT OF EXPERIENCE

At this point I would like to make a first attempt to relate the phenomenological conception of

<sup>5</sup> It was a happy surprise and confirmation of my thinking on this matter when Professor C. I. Lewis, in an unrelated letter of September 10, 1954, pointed out to me that "any directly discoverable item is found" and that "the given" in his writings was to be understood in the sense of "the sensuous found." I would not be prepared to replace the concept of the given completely by "the found." It seems to me rather that there are two types of the experienced, the given and the found, corresponding to two kinds of encounters with the world, the giving and finding experience. Yet the found in its distinctive nature has been comparatively neglected, although it would seem to be much more significant epistemologically.

experience to a concept which has become prominent in the discussion of non-phenomenological empiricism—that of the context of experience.

Now I must frankly confess that thus far I have not yet come across a clean-cut definition or description of this term in the contextualist literature. I gather therefore that its basic meaning has to be derived from common usage. Here I understand a "context" to mean the correlate of a "text" which is surrounded by the context. Or are we to understand that the text is not only the excluded core of a surrounding context but "part and pattern" of it? This would of course already be a widening of the literal meaning.

What, then, according to the contextualist empiricist, is the relation between an experience and its context? Is the context merely the outside frame for the "text" of experience? Or is it the whole, of which experience forms an integral part? If the former, then the context is at each given moment transempirical. But even if the latter is true, at least part of the context reaches beyond immediate experience.

Once these questions are cleared up, we can and must raise the further question: How do we know about the context? The question is particularly urgent for the transempirical concept of the context.

Now phenomenology, and particularly the phenomenology of constitution, should be in a position to tell us how the phenomenon of the context presents itself. Yet, actually, the term "context" does not often appear in phenomenological accounts of experience, and if so, only incidentally.<sup>6</sup> But this does not mean that the thing meant is absent. The terms that occur most frequently are "horizon," "field," "*Umwelt*" (the best English equivalent is probably "circumambience"), or, in the most comprehensive sense, "world." Obviously these expressions would cover only the transempirical sense of the contextualist's context. How, then, is it constituted for the phenomenologist?

Suppose we focus first on the thematic core or "text" of a configuration anywhere in this room—say, the table in front of us. In perceiving it as such, we are already aware of the concentric belts of decreasing clarity that surround it. These continuous belts terminate in a marginal fringe which melts away into the open blankness of a *terra incognita*. Now, while keeping the focus of our

primary attention on the perceptual core, let us try to watch and describe what connects it with the areas of diminishing attentional illumination. There is considerable variety in this regard in different perceptual fields. But there is probably something typical about the situation in the visual field. Here we may notice what may be called leads or tracks which run from the thematic field into the marginal belts and beyond. Outlines of objects and shadings may serve as illustrations. Now these tracks carry, as it were, references, or anticipatory "intentions" to the field beyond, foreshadowing not only what may be found in adjacent areas but also the perspectives in which they will be presented. They are decreasingly definite for the lateral aspects of an object we confront, and quite indefinite, though rarely absent, as regards its back.

What is thus vaguely presented or adumbrated as surrounding the focus of our experience constitutes the first and necessarily inadequate appearance of the field or context. Nevertheless, this is the original first-hand way in which we know about fields and contexts as such. Still, no isolated act of experience gives us adequate knowledge about our world. Only the further progress of our experience can orient us in this gradually constituted field. Its constitution takes place when we follow up the anticipatory references, making their vague content "thematic," or convert it into the "text" of our attention, while the previous core becomes marginal context and finally drops off into the background. From such a second step we may then proceed to check on other anticipations of our original intentions or on those which have newly arisen during our second step. It is in this manner that the context, first presented only vaguely and peripherally, constitutes itself into a firmer framework on an equal level with the first thematic text. But only when in the process of a sustained, but by no means rigidly patterned, sequences of confirmations, infirmations, and revisions it has constituted itself with a reasonable amount of clarity and stability, can the context claim equal rights with the text. And never can it achieve priority over it. For even after the constitution of the context, it is not only the context which determines the meaning of the text, it is also the text which supplies new meaning to the context. There is equality in interaction.

Nothing of what I have outlined here should be construed as an attempt to discredit or even to

<sup>6</sup> For a phenomenological discussion of context (or "thematic field") I can now refer the reader to Aron Gurwitsch, *The Field of Consciousness* (Pittsburgh, Duquesne University Press, 1964), especially pp. 318 ff.

question the pertinence of the concept of context, or the validity of contextualism. It is merely meant as an attempt to show that the concept of "context" is in need of further experiential analysis. A phenomenology of "text" and "context," which would thoroughly explore and describe their constitution in consciousness, may be one way of buttressing this promising development in latter-day empiricism.

#### VI. ON THE METAPHYSICAL SIGNIFICANCE OF A PHENOMENOLOGY OF EXPERIENCE

In conclusion I would like to discuss, however briefly, the question: What is the bearing of such a phenomenological analysis of experience upon metaphysics? The relation between phenomenology and metaphysics in general is actually one of the most open and controversial issues among living phenomenologists. It should also be mentioned that Husserl, who in the beginning phases of his phenomenology vetoed all speculative metaphysics, eventually, in his *Cartesian Meditations*, presented his idealistic monadology under the title of metaphysics.

All I intend to do in the present context is to outline what kind of connections might be found between the preceding analysis of experience and a conception of metaphysics which I have to define dogmatically. Be it understood, then, that by a metaphysical statement I mean here any statement which claims to reveal the fundamental structure of the world as a whole beyond mere appearance.

Is there anything metaphysical about the preceding study of experience in this sense? Two aspects in particular need discussion before an answer to these general questions can be given:

(1) Is the preceding analysis free from metaphysical assumptions?

(2) Does it add anything to the stock of metaphysical insights, if any?

In answer to the first question I plead: not guilty. Phenomenology suspends all belief in a reality beyond consciousness. This applies also to beliefs such as those in the reality of possible external causes of our experiences, whose truth sensationalism naively presupposes. Specifically, phenomenology does not commit itself initially to any belief in organisms as described by such sciences as anatomy, neurology, or physiology, or in an environment as described by the physical sciences. Only when the results of these sciences have been phenomenologically scrutinized and buttressed can phenomenology incorporate them.

This does not mean that phenomenology has no presuppositions whatsoever. No methodical examination can begin from the proverbial scratch. At least it will have to use such tools as logic and language. But these non-metaphysical presuppositions are not meant to be beyond all examination. They are always subject to phenomenological testing and retesting and in this sense accepted on approval only. I do not believe that in this respect phenomenology is any better or worse off than any other critical philosophy.

So much about the metaphysical assumptions of a phenomenological analysis of experience. What about the ultimate metaphysical significance of its findings? Is phenomenology the necessary and sufficient presupposition of any future metaphysics built upon experience? It would be preposterous to make such a claim at this time.

Can it at least promise a metaphysics of experience, i.e., a final appraisal of the essence of experience and of its place in the real world? This too would be premature. All that it can attempt is a clarification of the essential structure of experience. To decide whether or not such experience occurs in actual life is not its business *qua* phenomenology. Nevertheless this much may be claimed: The act of experiencing is part and parcel of the indubitable sphere of Husserl's absolute consciousness. Hence phenomenology can supply us with metaphysical knowledge about this one part of the universe.

The situation is much more complicated with regard to what is experienced, i.e., the object of our experiencing. Here we must remember that at least in its initial stage phenomenology suspends all claims to reality for everything that transcends the phenomena of consciousness. Its original stance is therefore one of metaphysical neutrality. The question is whether in its further development it can pass beyond it.

The following answer may serve at least as a hypothesis:

As far as the "what," the content of the experienced, is concerned, experience in its synthetic progress can supply increasing completeness, consistency, and corroboration. But it can never be exhaustive. All our knowledge of the content of the experienced world can therefore be only relative and subject to correction. If one wants to call the ultimate knowledge so achieved metaphysical, such metaphysics can never be absolute; at best it can have probability value.

As far as the "that," the being of this experi-

enced world, is concerned, phenomenology is in an even more precarious position. Husserl himself developed a peculiar transcendental idealism according to which the transcendent world owes its being to the constituting acts of the pure ego. Yet few other phenomenologists have followed him to this radical conclusion. Some have tried to formulate a phenomenological realism, whereas the French existentialist phenomenologists seem to be advocating an intermediate solution according to which existence and the world form a reciprocal system. Obviously in this matter phenomenology has not yet been able to reach a unanimous verdict. And frankly I do not see how it could at all, as long as it sticks to its original objective of exploring the phenomena *qua* phenomena regardless of whether they are real or not. Any decision that goes beyond this is, strictly speaking, transphenomenological. To be sure, even such a decision need not be an arbitrary decision. It can be based on good reasons, including phenomenological reasons. But these can never be compelling reasons.

*Washington University*

Personally I believe that a philosopher is entitled to make such metaphysical decisions. And I am not sufficiently concerned about the purity of the phenomenological faith to abstain from them in the name of the phenomenological suspension of belief. But I also wish there were more purists who would not dilute phenomenology for the sake of metaphysical prizes.

My all too sweeping conclusion, then, is that pure phenomenology as such is not a sufficient foundation for any future metaphysics of being as a whole. It can only supply a metaphysics of consciousness, including that of the experiencing consciousness, as one of its constituents. Besides, it may supply relevant, though not conclusive evidence for transconscious reality. Once phenomenology has done more of a solid job in screening this evidence, it may become the indispensable foundation for metaphysics. But today phenomenology had better concentrate on this foundation until it can support a superstructure that will not prove top-heavy.

## V. CORRIGENDA

### VOLUME I (1964)

JUDITH JARVIS THOMSON: *Private Languages* (pp. 20-31).

Page 20, col. 1, line 7: Replace "trite" by "tired"

Page 20, col. 2, lines 15-18: Replace "Were one to . . . in both cases," by "Which, if you look at the nature of the arguments which have actually been produced *for* Malcolm's thesis, would be to do exactly the same thing in both cases,"

Page 20, col. 2, line 34: Replace "this" by "thus"

Page 22, col. 1, line 13: Omit "in"

Page 22, col. 1, Part IV line 2: Omit "in"

Page 23, col. 1, line 12: Replace "if" by "it" and omit "were what"

Page 23, col. 2, Part V line 13: Replace "need not" by "must not"

Page 25, col. 1, line 17: Omit "one"

Page 25, col. 1, line 34: Replace "thought" by "decided"

Page 26 (throughout): Replace "identify" by "pick out"

Page 28, col. 1, line 41: Replace "what" by "that"

BRIAN MEDLIN: *The Unexpected Examination* (pp. 66-72).

Page 70, col. 1, line 20: Delete "we obtain"

Page 70, col. 1, line 21: Delete "although"